APPLICATIONS OF RESEARCH

ON HUMAN DECISIONMAKING

A symposium held at
AMES RESEARCH CENTER
MOFFETT FIELD, CALIFORNIA
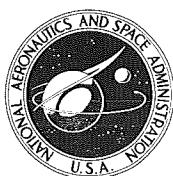January 31–February 2, 1968

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

APPLICATIONS OF RESEARCH

ON HUMAN DECISIONMAKING

A symposium held at NASA Ames Research Center
January 31–February 2, 1968, and sponsored by the
Human Performance Branch of the Biotechnology Division

Edited by
R. MARK PATTON AND TRIEVE A. TANNER, JR.,
NASA Ames Research Center, and JOSEPH MARKOWITZ
AND JOHN A. SWETS, *Bolt, Beranek & Newman, Inc.*

Preface

This symposium on the Applications of Research on Human Decision-making was held at the Ames Research Center and was sponsored by the Human Performance Branch of the Biotechnology Division. This symposium and many of the studies reported are part of NASA's Human Factors Systems Program, directed by Walton L. Jones, M.D., and the Electronics and Controls Program, directed by Frank J. Sullivan.

These proceedings reflect some of the efforts underway to close the gap between basic research and its applications.

Foreword

In recent years, there has been a general trend toward increased complexity in man-machine systems. Examples of particular interest to the National Aeronautics and Space Administration include large and fast aircraft, manned spacecraft, and elaborate ground-based facilities for the control of space missions. Typically, such systems employ man primarily as a systems manager, with many requirements placed upon him for complex decisionmaking. Such emphasis makes it likely that any human error in the operation of such systems will be one involving decisionmaking.

Laboratory workers in the behavioral sciences have become increasingly concerned with studying complex behavioral processes. Early laboratories stressed investigations of such relatively simple processes as sensory discriminations and reaction times. There has been an increasing interest over the past few years in complex information processing and decisionmaking, and a substantial body of information has been developed. This research has developed not only from concerns with hardware systems but also from problems of education, social interaction, and so forth. Thus there is a growing body of laboratory data available for application to NASA concerns.

The Human Performance Branch of the NASA Ames Research Center has been deeply involved in this type of research for several years. This work has been supported by the Biotechnology Division and by the Electronics and Control Division of the NASA Office of Advanced Research and Technology. As elements of a mission-oriented agency, we are particularly concerned with the application of this information to operational problems.

Laboratory research in the behavioral sciences employs simplified analogs to the "real world." Such simplification involves a risk that essential elements may be missing, making comparisons between the laboratory and the operational situation invalid. The development of valid methods permitting transitions between the two domains depends upon a person (or a team) having substantial familiarity with both. Such people, and such teams, are rare.

It was the purpose of the conference on Applications of Research on Human Decisionmaking to bring together representatives of the two domains in order to make each group more aware of the concerns, requirements, and capabilities of the other.

R. MARK PATTON

v

Contents

Introduction

JOHN A. SWETS

Bolt, Beranek & Newman, Inc.

This conference is concerned with the applications of psychological research to the operational problems that NASA faces. Although the theme of this conference is "human decisionmaking," human information processing will also be discussed. Some of the speakers will emphasize the parts of their research that seem to them to be applied or to have potential applications. There will be variability, of course, within and across speakers. In fact, for some of the research that is going on, it is not entirely clear that there are immediate applications. Outside discussants have been invited partly to ease this confrontation of researchers and people who know about real problems and to facilitate the conversation between them. The discussants know about the research and have some background with the kinds of problems to which we want to make applications; they have also done some research themselves.

Decisions are ubiquitous. If we sometimes have misgivings about how effectively psychological research is applied, at least we can feel confident here that we have a topic that is significant. We can hardly avoid decisions, and, understandably, the study of decisionmaking is very complex. We are going to consider decisions in several contexts: the pilot's task, sensory processes, motor processes, probability judgments, and so forth. There are, however, a variety of kinds of research on human decisionmaking that we will not discuss, and it will be well to keep that in mind as we welcome the first speaker.

Decisionmaking in the Manual Control of Aerospace Vehicles

GEORGE A. RATHERT, JR.

NASA Ames Research Center

These remarks concern the role of decision-making in the operation and control of piloted aerospace vehicles. Specifically, we will review selected portions of the engineering research used to develop manual control systems to point out areas where the decision-making process is a significant problem in establishing satisfactory dynamic behavior of the pilot-control system-vehicle combination. The motive is to interest you in these problems and to describe them with sufficient clarity to permit you to evaluate the possible contribution of your own studies. Obviously, many different decisions occur at a formidable rate in any given flight; but I have limited consideration to problems for which I feel there is a reasonable chance of fruitful interaction between our disciplines.

With this in mind, we will primarily consider problems actively studied on piloted simulators, for a simple but very significant reason. Generally speaking, if engineers have a problem on a simulator, an attractive experimental setup exists in which you can participate by asking. By definition, a successful simulation presents a specific problem or decision to be made. The presentation is often limited to the factors relevant to that decision; it has reliable presentation of cues, unlimited reproduction of the experiment, and uses sophisticated subjects, test pilots, and astronauts with established proficiencies of a caliber that very often are not accessible to experimenters in any other way. In other words, it is a running start on a pertinent experimental design that might satisfy even a human-performance expert.

This sounds like we are trying "to establish a more productive relationship between mission-oriented simulation and human-performance research," a phrase taken from a paper by Knowles (ref. 1). This paper fits the theme of our conference, and I would like to relate some of Knowles' ideas to the situation, at least at Ames.

There is a continuous spectrum between aerospace simulation and human-performance research. At one extreme is the one-shot mission-oriented task simulator in which our chief pilot, George Cooper, makes a yes-no decision on the feasibility of a specific pilot task; at the other extreme are the signal-recognition experiments in Trieve Tanner's laboratory. We are all aware of the friendly bickering that points out the deficiencies of either approach. The task simulator yields little data that is valid for a general case or directly applicable to predicting the performance of a new and a different task; and Dr. Tanner's model situations will not tell us if Mr. Cooper can actually land the airplane.

There is, in the middle of this spectrum, what Dr. Obermayer (ref. 2) has called the forgotten man—the systems engineer who, on the basis of past findings at both ends of the spectrum, must predict future designs. His principal task is ultimately to develop and express a successful methodology for manned-systems design. The point is that there are times when experts should move in

from both ends and help him out, even if the urgent problem on the task simulator has an overbearing priority and if the theoretical model is not quite yet as tidy as one would like it. I am not naive enough to suggest a series of futile piggyback experiments, but I think we can manage our large, sophisticated simulation complexes so that sufficient work is accomplished to gain three objectives: in Knowles' terms, the feasibility demonstration, valid general empirical data, and adequate model or theory verification. However, we are not going to be able to accomplish these objectives without adequate participation by people who know how to design and conduct human experiments. The balance of this paper, then, is a review of those areas involving the decisionmaking process in which I would like to nudge the human-performance experts a little bit.

The first problem to be faced is this: how to classify situations involving decisionmaking so as to present them in some orderly way. I am a devout coward about categorizing decisions with behavioral vocabulary; therefore, the problems will be grouped on the basis of what the pilot is trying to do. There are four major groups:

Group I: Immediate control of attitude about the flight path

Group II: Short-term (up to 1 minute) control of the flight path

Group III: Long-term control of the flight path

Group IV: Overall mission or systems management

In an actual mission, of course, these are all occurring nearly simultaneously in real time. As we discuss them, you will note a steady growth of the relative role of decisionmaking.

Group I problems characteristically place the pilot in the role of an attitude sensor and dynamic neuromuscular controller who must stabilize unwanted motions created by the interactions of the control system, the vehicle, and the environment. Figure 1 rep-
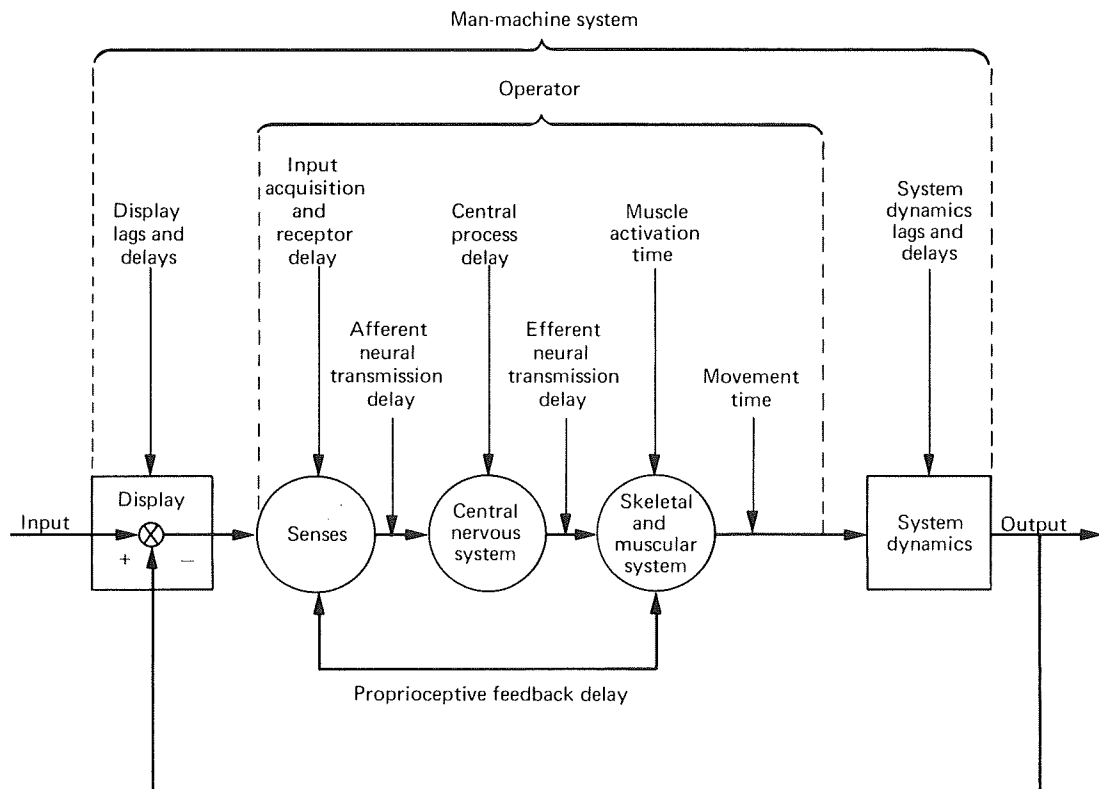

FIGURE 1.—Manual control process, developed by Wargo (ref. 3).

resents this control process in a form developed by Wargo (ref. 3). I like it because it uses the word "delay" so many times. We can summarize figure 1 by saying that the pilot goes through four major processes—perception, decision, response, and feedback—in what typically must be a fast-responding, stable closed loop with the delays reduced to the human minimums cited in the literature.
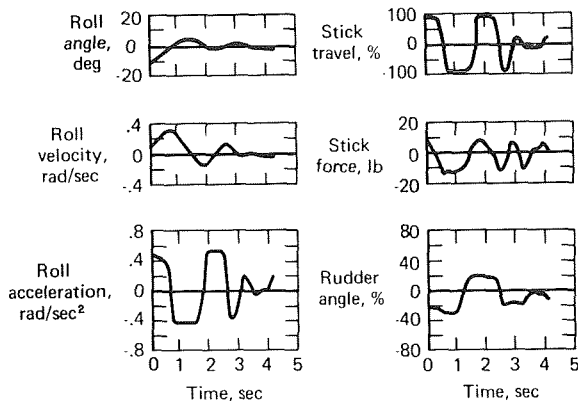

FIGURE 2.—Typical attitude control of pilot in V/STOL airplane.

In many tasks in group I, the pilot is required to develop almost a reflex behavior. As figure 2 shows, hovering a V/STOL airplane is an example. Here, decisionmaking is unimportant and, of our four processes, perception and response are the important dynamic parameters; they are the only significant delays. Notice on the time history the characteristic bang-bang response of full-control travel in response to a perceived disturbance. Within reasonable physical limits, this is the kind of set behavior that the pilot can perform while carrying on a conversation or thinking about other problems. In this situation, the design engineer's task is to set numbers on the pilot's perception and neuromuscular response, and then select the control system and vehicle dynamic performance (the flying qualities) so that the overall system is sufficiently fast and stable to cope with an adequate range of environmental conditions. He can check out his solution on a piloted simulator or variable-stability airplane.

Where decisionmaking is not a problem, I am going to follow through and very briefly illustrate the successful use of flying-qualities specifications and human pilot models. It is a digression in a sense, but I want to give you a feel for the procedure that we are groping for in the areas where the decision delay is vital. Figure 3 is a typical flying-qualities chart (ref. 4), which shows two of the many dynamic parameters of the vehicle to be controlled: longitudinal short-period damping and stability. Figure 3 simply illustrates that, given perception and response delays typical of a normal pilot, various combinations of damping and stability match to give an acceptable overall response, whereas other combinations would be unsatisfactory, and would require the change of some dynamic parameter.
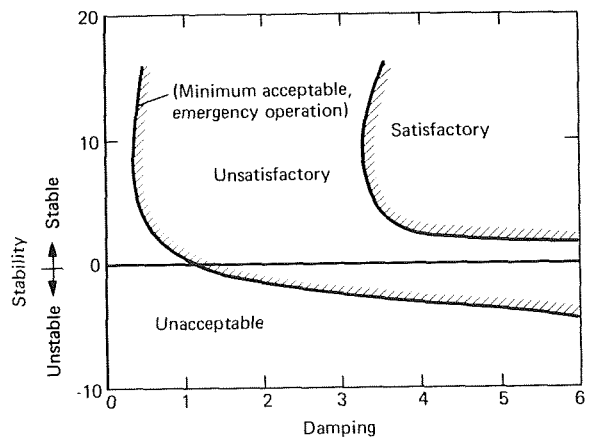
FIGURE 3.—Typical flying qualities analysis that emphasizes longitudinal short-period damping and stability, two dynamic parameters of the vehicle to be controlled.

Figure 4 illustrates the use of mathematical pilot models. This is important to me because we are continually being challenged as to why we fool with these models. Do we not know about the infinite variability of the human operator, his marvelous adaptability to higher order control techniques, discrete control modes, and so forth? Of course we know about these things, but even a greatly simplified model gives useful
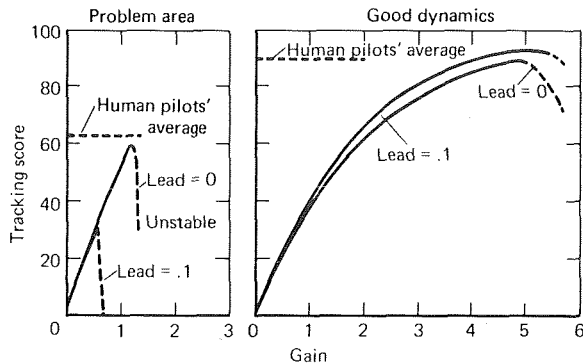
FIGURE 4.—A typical use of mathematical pilot models.

FIGURE 5.—A time history of a typical pitchup problem.

insight. Figure 4 shows a simple tracking task performed with two different control systems (ref. 5). The real human pilot did quite well with one system, but his performance with the other is obviously inferior. We applied a well-known, simple mathematical model with five dynamic parameters. By manipulating two of these, the model pilot gain (stick motion per error) and lead (or anticipation), we were able to derive the actual performance in both cases. With the superior system, the real pilot could have used a wide choice of response to error and did not have to worry about supplying anticipation at all to get the desired performance. With the inferior system, the choice of gain is quite critical and the pilot must supply lead. The point is that the model is not used for quantitative behavior prediction; it is used to tell us before the systems are even built that the kind of system that would require this control behavior is inferior—particularly so in a situation where we are going to make other intellectual demands (e.g., decisions) on the pilot.

To get back to our main theme, in group I, the reflex behavior is not adequate; decisionmaking joins perception and response and takes enough time to impair seriously the dynamic performance of the system. Various classes of these will be described by single examples.

Group I consists of various aerodynamic instabilities of the basic airplane and is illustrated in figure 5 by a time history of the

pitchup problem (ref. 6). When some airplanes reach a certain angle of attack, their stability characteristics begin to change because of wingstall or downwash phenomena. Instead of tending to straighten out by themselves, they begin to diverge and, if left alone, will pitch on up to the stall. Figure 5 shows a pilot in a normal windup turn with the g-factor increasing steadily. At some point, the pilot perceives that the rate of increase is exceeding what he is calling for by the controls and he must execute a recovery and push forward on the stick to prevent excessive g-loads or a stall. What the pilot senses, incidentally, is an uncontrolled-for pitching acceleration developing past his perception threshold of about 0.15 rad/sec².

However, a simple reflex recovery action will not do. You can see on the tail-load trace that, depending on the abruptness of recovery, very large maneuvering tail loads can be developed. Whether these will exceed the design limits depends on speed, altitude, and rate of entry into the maneuver. The pilot's task is far from a simple reflex. He must perceive the motion, decide its true source among many possibilities, and weigh the hazards of delayed or excessive response in

the circumstances of the moment. Quickening the perception or response is not the answer; the pilot needs aid in the decision process. Incidentally, in lieu of being able to aid or solve the decision problem, the alternative for this specific problem has been to install elaborate mechanical systems to sense the problem and push the stick forward at programed rates depending on the flight circumstances. To help the pilot in this way, a high penalty is paid in terms of weight, reliability, and maintainability.
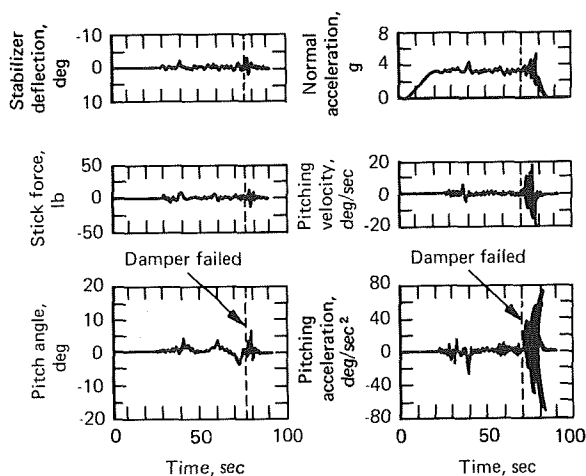
FIGURE 6.—Sudden failure of a pitch damper in the control system.

The next problem area, with a different type of decision involved, is sudden control or structural failures that are not surely catastrophic. Figure 6 shows, as a typical example, the sudden failure of a pitch damper in the control system (ref. 7). I deliberately chose an example in which the pilot is on the verge of adapting and maintaining control but finally loses it. Obviously, depending on the circumstances and the particular failure, the pilot sometimes can maintain control and at other times faces a hopeless task. Figure 7 shows the predicted behavior of a fighter airplane when the pitch damper fails at low altitudes and the pilot cannot adapt. Note the divergent buildup in g-forces to 16 g in less than 2 seconds. Obviously, the motion must be perceived, the

source identified, and the correct decision made (eject) in less than 2 seconds to preserve a reasonable chance for a safe ejection. This is just one example of a situation with a very short time constant. There are many aerodynamicists and systems people devoted to eliminating these problems entirely, but there are still too many cases of "sticking with the ship" past the point where an aided or improved decision process would have triggered a safe ejection.

As another example, consider a deep stall and spin recovery of, say, an F4H. Obviously, the airplane gets into an aerodynamic configuration that it cannot get out of; yet the pilot thinks it can and stays with the plane too long. These kinds of decisions are made in an environment where I think the pilot needs help.

A third example in which decision must be interjected between perception and response is termed turbulent-air upset. Figure 8 is a time history of g-force and altitude

FIGURE 7.—Predicted effect of damper failure on the flight path.

FIGURE 8.—Time history of a typical upset in turbulent air.

from a commercial airliner flight. In this case, the pilot perceives a departure from his desired flight attitude induced by turbulence and chooses to stabilize his attitude by paying attention to the wrong cue, in this case, airspeed. Without realizing the implicatio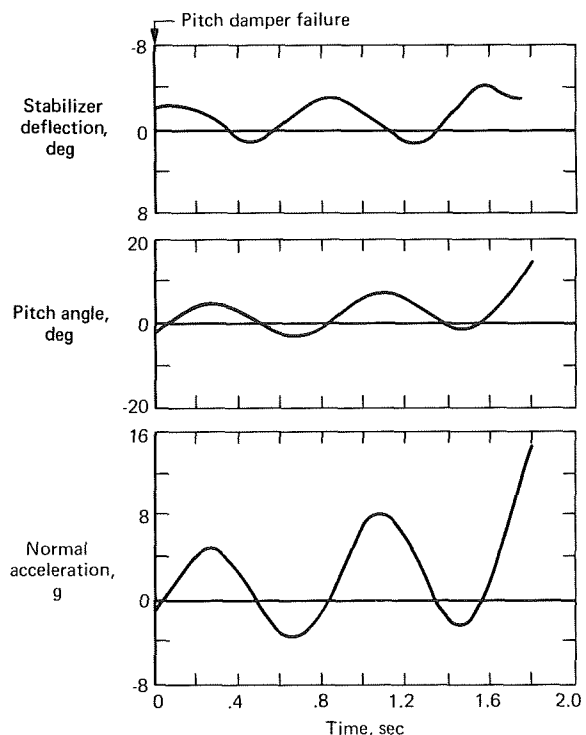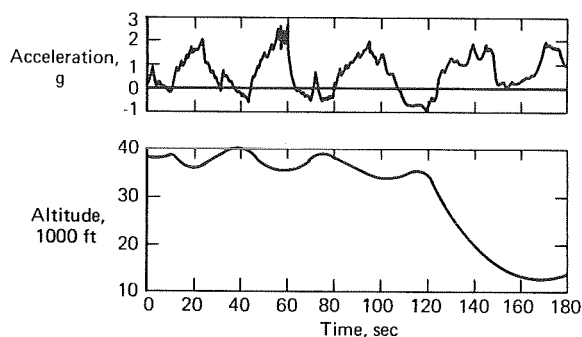ns of this decision, he persists and builds up a dynamic oscillation in angle of attack [not shown], altitude, and g-force. This oscillation persists until minutes later, when recovery can only be effected by a dive, resulting in a loss of thousands of feet of altitude, which he was fortunate enough to have. The various aerodynamic considerations forcing this choice of recovery are too lengthy to detail here, but the necessity for training and reinforcing the correct decision process in the pilot is obvious.

A fourth example of an immediate attitude control problem in which the decision delay becomes critical is disorientation in flight. Figure 9 is a time history of altitude and acceleration in a real flight from take-off. We have all at least read about laboratory experiments on sensory deprivation in closed rooms, but this incident happened to the captain of a commercial airliner with nearly 100 passengers on board. He took off, climbed through heavy turbulence, and entered clouds. The turbulence built up to quite high oscillatory levels, including negative 1 g, with all kinds of sensory cues and instrument readings. This particular airplane has windows above the captain's head and, at this point, he perceived a pattern of

lights through the overhead windows. He had no trouble interpreting that cue and, in a superb piece of flying, quickly got the ship in free fall on its side and recovered at about 1200 feet. There are two problem areas here: selecting and presenting unequivocal information to permit a quick decision that disorientation is occurring, and guidance in deciding on the recovery technique.

One common factor that complicates many of these problems is false cues or uncertainty about the source of the motions that the pilot perceives. Pitching acceleration was the initial cue detected in each of our example problems; however, without an orderly decision process, it would obviously be difficult to identify at the threshold level whether the pitching acceleration was due to aerodynamic instability, vehicle systems failure, turbulence, or inattention. With the newer generation of long, slender, flexible aircraft, there is also an ambiguity of sign. Figure 10 shows the situation of the pilot in the forward cockpit. A hard-over actuator failure, causing the basic vehicle to pitch up, could actually give the pilot the sensation of pitching nose down; not a good start for his instinctive emergency recovery response.

To summarize the group I problems briefly, we have examined several types of problems in the immediate control of attitude about the flight path where inserting a decision process must augment the reflex type of pilot behavior, perception, and re-
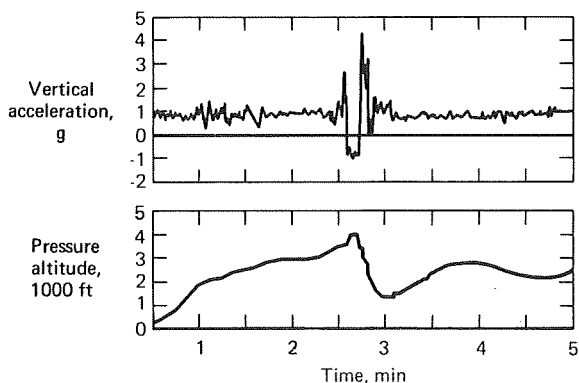

FIGURE 9.—Time history of altitude and acceleration, showing typical disorientation and recovery.

Initial g loading
at pilot station

Actuator
hard over

Rigid body rotation

Initial displacement
of body centerline due
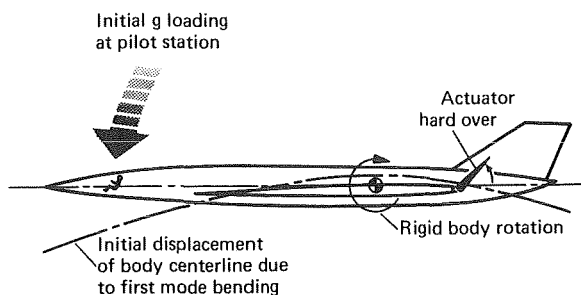to first mode bending

FIGURE 10.—Structural flexibility—the situation of the pilot in the forward cockpit during a hard-over actuator failure.

sponse. The motion cue must be analyzed to decide its true source. In some cases, delayed response was hazardous; in other cases, excessive response was hazardous. In all cases, much effort can be wasted on facilitating the pilot's perception process or assisting the speed of his control-motion response if what he needs is help in the decisionmaking process.

Next I will consider group II problems: short-term control of the flight path. In group I, the pilot had to control his attitude but, within reasonable limits, really did not care about his flight path. In group II, he will have all his group I problems, with the additional task requiring precision control of the flight path as well. In other words, the pilot is still an immediate attitude controller and, in addition, he is a fast-acting predictor of the immediate flight path. Here the decisionmaking process is nearly always as important as perception and response.

The first example considered is low-altitude high-speed flight or terrain following (ref. 8). Figure 11 is a pictorial view of a typical flight path as a function of elapsed

time. Figure 12 shows typical displays that the pilot uses to perform this task on manual control. In the display at upper right, the pilot sees the relative height of the terrain, 10 and 5 seconds ahead, the terrain below, and a heading indicator. There are other predictive display games, such as memory dots for terrain maxima and so forth, but the basic problem is clearly one of understanding and properly aiding the decisionmaking process. Incidentally, the environment can be quite rugged. Figure 13 shows a time history of the normal acceleration in a flexible airplane. Both actual and simulated missions with this task have been successfully flown for 1 to 3 hours.

The next example problem in group II is weapon-systems operation. I will use older, hence unclassified, examples, but the principles are the same. Figure 14 illustrates the pilot task in tracking with an open disturbed-reticle sight (ref. 9), a task that is regaining its earlier prominence. The pilot controls the gun line or airplane axis with his usual group I attitude-control problems; however, here he also must control the flight path to keep the line of sight on the target. The line of sight is not fixed to the airplane but moves with respect to it, by the dynamic equation shown in the box. Hence, the pilot must operate the gun line through the airplane control and environment dynamics so that, when moved by the gunsight dynamics, the line of sight stays steadily on the target. This is almost a classic example of divided attention: simultaneously solving two sets of dynamic responses.

The effects of this task on flight techniques and training are quite marked. When we first started to study the effects of airplane

(Simulated bending mode)

Altitude, 1000 ft

Time, sec
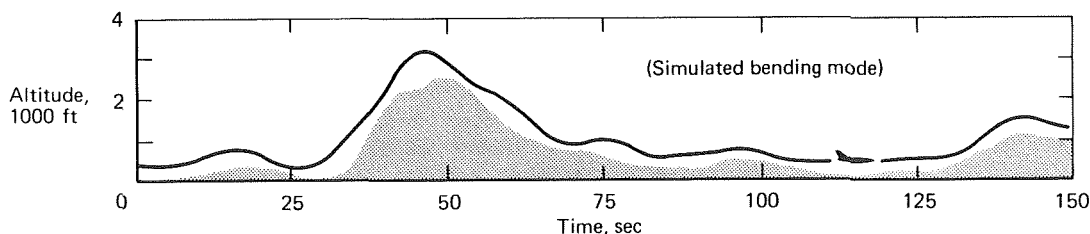
FIGURE 11.—Typical flight path as a function of elapsed time.

Display C
Angles to terrain ahead scaled 6°/cm
terrain below 333 ft/cm
using horizon center as zero reference

Heading indicator (off course)
Terrain 10 sec ahead
Terrain, 5 sec ahead
Terrain below

Display D
Terrain ahead shown as
relative heights, 333 ft/cm

Display E
Height memory dot (maxima of
terrain 10 sec ahead) added

Height memory dot
Miniature airplane

Display F
Heights rescaled to 250 ft/cm
pitch angle = 2.2°/cm

Horizon
(right bank,
nose down)

FIGURE 12.—Display nodes used to perform task
(illustrated in fig. 11) on manual control.

FIGURE 13.—Time history of normal acceleration in
a flexible airplane.

(Simulated bending mode)

Simulator cockpit
acceleration, g

Time, sec

Line of sight
Tracking line
Gun line
Flight path
Relative wind

Lead angle transfer fn:

$$P = \frac{(S/k)}{1 + (c/k)D} \dot{L}$$
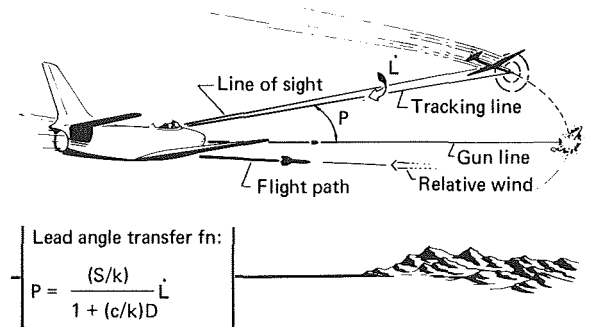
FIGURE 14.—The pilot task in tracking with an open
disturbed-reticle sight.

dynamics on tracking ability, even with fixed sights, we noticed that the Ames test pilots with Navy backgrounds handled the airplanes relatively harshly and used very quick and high-frequency control motions to keep on target. The ex-Air Force pilots, on the other hand, flew more gently and smoothly and made corrections relatively slowly. Investigation showed the cause was right in this box: the Navy had a very stiff, close-coupled disturbed-reticle sight that stayed close to the airplane axis at the expense of being slower to respond and show the correct lead. The standard Air Force sight had a very high dynamic response to show the correct lead more quickly, but was very easily disturbed and had to be handled gently to prevent oscillations. Both of these represent reasonable compromise design solutions, but result in markedly different piloting techniques due to the forcing effect on the pilot's decisionmaking process in short-term control of the flight path.

Another problem that should be mentioned in group II is collision avoidance. Although perception is the mandatory first step, the pilot's decisionmaking process is also becoming critical. Increasing speeds, larger and less maneuverable airplanes, and increasing

traffic are all negative factors. Here we see considerable promise in automatic assistance; however, a feasible and economic solution must be based on an understanding of the human pilot's task and an analysis of the help he actually needs.

To summarize the group II problems, the relative importance of decisionmaking has increased and there is obviously room for applying predictive displays, controls, and warning devices to augment the pilot and unload his decisionmaking capability. Considerable care and understanding of the pilot's actual behavior must be used, however, to select the optimum form of assistance that the pilot really needs.

Group III problems—long-term control of the flight path—differ from those of group II in that the speed of control response may become less critical, but the tasks become more complex. The relative role of decisionmaking (or, perhaps more properly at this level, information processing) is much greater.

The first problem to be considered is that of weapon-systems operation wherein the pilot's control task varies significantly with time. In the preceding (group II) example of a disturbed-reticle sight, the pilot was required to keep the reticle on target for some length of time, but the behavior pattern required was constant and did not vary through his attack. He was firing a string of bullets, which takes time, and the sight needed time on target to have a chance to compute the correct lead angle, but the control pattern did not change.

There are other weapon systems that need to be aimed precisely only at the instant of firing. To unload the pilot and free his attention for some of our other problems, it would be desirable to let him track rather loosely for most of the time and then have him tighten up and give his undivided attention and best performance at the critical instant of missile launch. Figure 15 shows a director display designed to force this decision (ref. 10). The pilot's task is to fly so that the larger circle in the displays surrounds the target. It is a large circle and is displaced
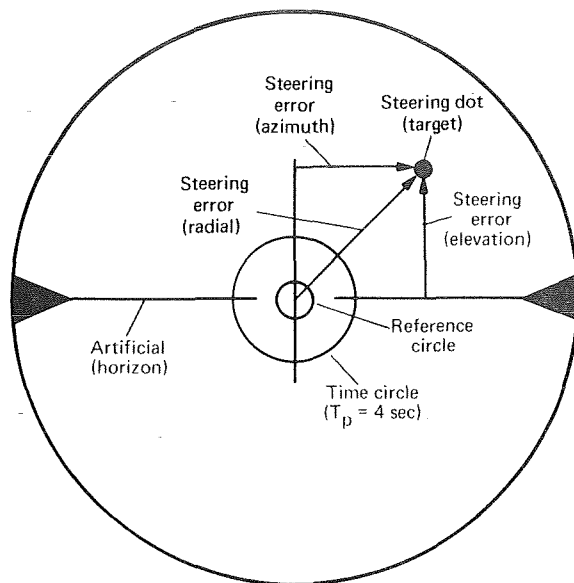
FIGURE 15. — Weapon-system operation — director display.

from the target dot by a dynamic response that is fairly stiff and not easy to upset. As we approach 4 seconds to go, however, the large circle starts to collapse to the size of the smaller one and the dynamic response increases to pinpoint the lead angle, making it more difficult to cope with and making it mandatory for the pilot to divert his attention and concentrate on this task.

Obviously, there are a number of other weapon systems that require divided attention, control of two sets of dynamics, and time-varying control precision, but it is interesting to observe how many new developments can be rationalized on the basis of unloading the pilot's basic decisionmaking process in the disturbed-reticle-sight task. We concentrate the armament load to reduce the performance firing time, we put one of the two sets of dynamics in an automatic missile, we signal the pilot-control-gain changes with director-display games, we increase the lethal radius of the missile to make the whole task performance less critical, and so on.

The next example of a manual-control task for long-term control of the flight path is navigation. At this point, we must comment

that in group III we have arrived at a level of task complexity in which the overall systems designer has a real choice between either manual or automatic operation. In the problems I am now presenting, there are alternative and possibly better methods than manual control, but two points should be noted: the economic and feasible choice must be made in view of accurate knowledge of the manual-control capability; and, in most cases, a workable manual-control scheme will still be required for emergency backup.

The manual navigation task used as an example is for an emergency return from a lunar mission. One scheme that has been developed is the manual use (by the pilot) of a hand-held sextant (ref. 11). A simulator was used to determine the feasibility of this approach. The pilot in the moving-base capsule read specified angles between collimated targets in his field of view. The accuracy of the measurements was quite good—on the order of 10 seconds of arc standard deviation with adequate training and experience. The important point to us is that critical delay in the system was the decision by the pilot that a particular sighting was a valid measurement. Again, from an overall system point of view, the pilot's chief need for help was not in the manual measurement technique but in the decisionmaking process.

Probably the most prominent examples of problem areas in long-term control of the flight path are the takeoff and the landing. Particularly in a commercial jet transport where the takeoff seems to be one almost continuous decision tree, the decisionmaking process is almost more important than the manual-control performance, especially in emergencies. The takeoff decision tree is too lengthy to present here with clarity; I will recommend to those interested an excellent book by D. P. Davies, *Handling the Big Jets*. Davies is the chief pilot of the British Air Registration Board and he speaks volumes in the pilots' language.

To summarize the group III problems, long-term control of the flight path, we can discern two trends that have been increasing through groups I and II. The relative role

of the decisionmaking process in comparison with perception and response has become much greater. Also, as we accumulate the previous problems and add them together, we have arrived at the stage where it has been found necessary and feasible to solve some of the more complex problems by installing semiautomatic or completely automatic systems to unload the pilot's decisionmaking function by taking over part of his tasks.

However, we greatly value the human pilot's decisionmaking capability and, in a sense, much prefer using him this way to using him as a superadaptive servocontroller. This brings us to the problem of properly allocating functions between pilot and machine, which I want to stress as my final area for nudging the human performance experts. Figure 16, which is from a study by Serendipity Associates (ref. 12), shows a spectrum of possible approaches to a manned system. The upper solid line is a completely automatic approach; the lower broken line is a completely manual approach. Obviously, if there is to be a normal "ornery" human pilot in the system, and the task is on the level of the problems we have been considering, all the possible system-design alternatives are somewhere in between. In each of the problems we have discussed, the possibility of assisting the pilot varies from trivial to an automatic takeover. Determining the optimum allocation of function, from the standpoints of both the human pilot and the overall system performance, is a serious problem. Its successful solution greatly depends on accurate knowledge of the human decisionmaking process and its quantitative performance limits.

We have next the group IV problems: overall systems or mission management. This group has been noted here primarily for completeness of outline. Although the decisionmaking process is, perhaps, of paramount importance in this area, we at Ames have contributed little that I am competent to discuss, partly because in this group we tend to get away from the manual-control problems of a single pilot to those of man-
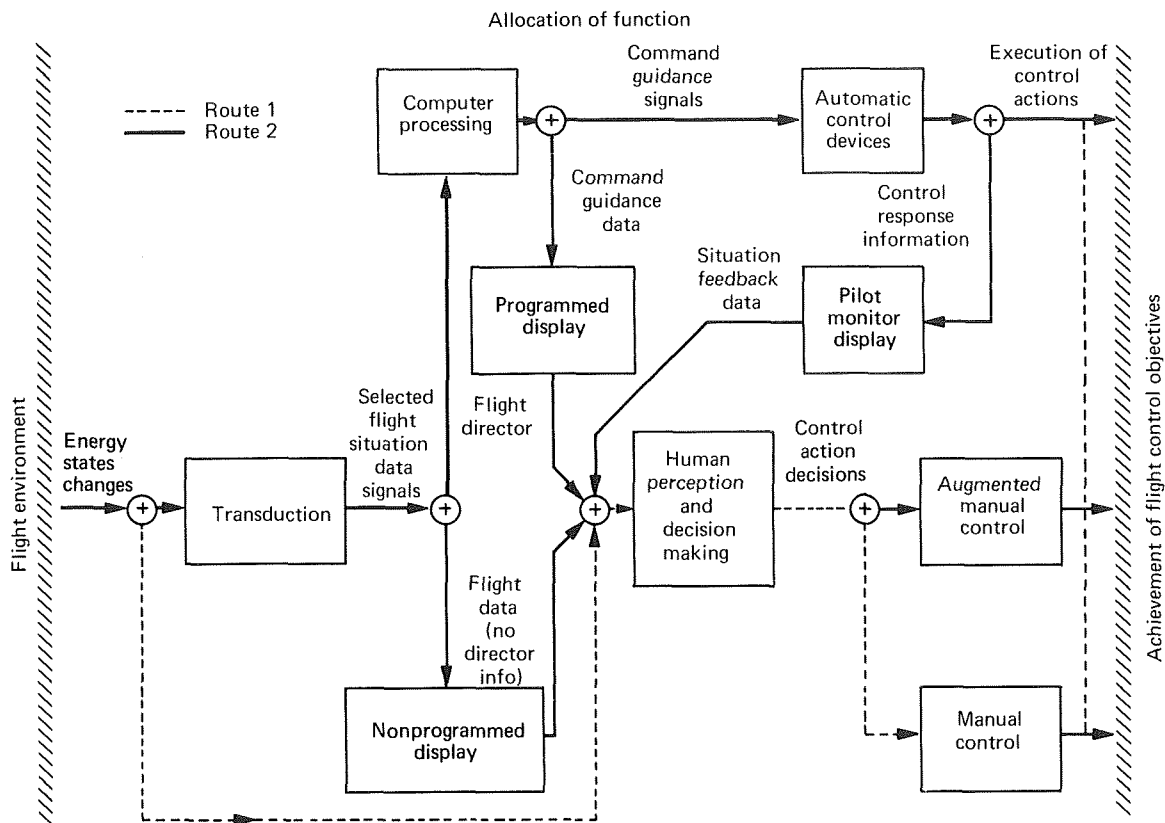
FIGURE 16.—The spectrum of possible approaches to a manned system.

aging more complex organizations. Harold Miller will describe some typical problems in this group in a succeeding paper. In a sense, group IV also needs less discussion from me because human-performance experts are already involved and the importance of the decisionmaking process is so obvious.

Figure 17 shows how all of the previous groups of problems coalesce into the structure of a real mission. Figure 17 is a time history of a pilot adapting to an attitude-control failure while simulating a manually controlled reentry. The simulation was done more than 10 years ago and many such studies have contributed to the manual-control mode that was available for astronaut Cooper's reentry. In what context, however, did the problem actually occur on Cooper's mission? (*Mercury Project Summary*, 1963). The failure occurred and was perceived and coped with immediately by

manual control of attitude as shown here. A mission-control decision tree isolated the cause and extent of systems failure. The informed Cooper then looked for a substitute
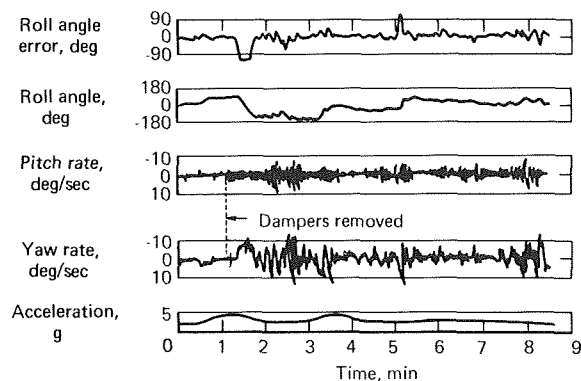
FIGURE 17.—Time history of a pilot adapting to an attitude-control failure while simulating a manually controlled reentry.

visual reference for manual flight-path control. He found one in the city lights of Shanghai and then continued immediate control of attitude and programed it to execute the desired long-term control of the flight path for a successful reentry. My final point is this: we can isolate these problems in our talks, our laboratory experiments, and theoretical models, but they seem to occur simultaneously in real-life situations.

DISCUSSION

JEROME I. ELKIND: You have said that decision problems are involved in some of the examples of failure in group I tasks; in other examples you said decision problems are not involved. I do not really see the distinction. It seems to me that one possible decision is, of course, ejecting; another is to adopt a different kind of control action.

GEORGE A. RATHERT, JR.: You must decide to adopt the different control mode. If you do adopt a different control mode, then you must evaluate whether you are being successful or not. It would be very nice if someone who knew enough about the situation would be able to test and evaluate very quickly whether or not the adaptation was successful.

How do you decide quickly enough whether or not the adaptation is successful? To do this requires prediction techniques of human behavior that I think would be pertinent to the problem, but the right type of people have not studied it. In other words, when the damper fails, the pilot immediately begins tracking. Something should be able to tell him right away whether or not he is beginning to cope with the problem. He is so busy in the reflex response motion that I do not think that he can make the decision properly. Perhaps something can measure what the pilot is doing and then tell him.

STEVEN E. BELSLEY: The basic problem is really to eliminate the occurrence of damper failure because the machine comes apart before the pilot has time to think about it.

RATHERT: People are trying to eliminate these problems entirely. The point is that they have not yet succeeded. Until they do I think that this is an area in which we can help the pilot.

REFERENCES

1. KNOWLES, W. B.: Aerospace Simulation and Human Performance Research. Human Factors, vol. 9, no. 2, 1967, pp. 149–159.
2. OBERMAYER, R. W.: Measurement Criteria in Man-Machine Systems Simulation. NASA CR–257, 1965.
3. WARGO, M. J.: Human Operator Response Speed, Frequency, and Flexibility: A Review and Analysis. Human Factors, vol. 9, no. 3, 1967, pp. 221–238.
4. RATHERT, G. A., JR.; CREER, B. Y.; AND SADOFF, M.: The Use of Piloted Flight Simulators in General Research. Presented at the AGARD Specialists' Meeting on Stability and Control (Belgium). AGARD Rept. no. 365, Apr. 10–14, 1961.
5. SADOFF, M.: The Effects of Longitudinal Control-System Dynamics on Pilot Opinion and Response Characteristics as Determined From Flight Tests and Ground Simulator Studies. NASA Mem. 10–1–58A, 1958.
6. SADOFF, M.: Summary of the Flight Conditions and Maneuvers in Which Maximum Wing and Tail Loads Were Experienced on a Swept-Wing Fighter Airplane. NACA RM A55A06, 1955.
7. WHITE, M. D.; SADOFF, M.; BRAY, R. S.; AND COOPER, G. E.: Assessment of Critical Problem Areas of the Supersonic Transport by Means of Piloted Simulators. Paper presented at 30th annual meeting, Inst. Aerospace Sci. (New York, N.Y.), Jan. 22–24, 1962.
8. SADOFF, M.; AND WEMPE, T.: Simulation of Pilot-Vehicle System for Low-Altitude, High-Speed Flight. Presented at AGARD Inter-Panel Specialists' Meeting on Low Altitude, High-Speed Flight (Paris), Oct. 20–23, 1964.
9. RATHERT, G. A., JR.; ABRAMOVITZ, M.; AND GADEBERG, B. L.: The Effects of Powered Controls and Fire-Control Systems on Tracking Accuracy. NACA RM A55D12a, 1955.
10. TURNER, H. L.; RATHERT, G. A., JR.; AND HEINLE, D. R.: Tracking Performance of a Swept-Wing Fighter with a Direction-Type Radar Fire-Control System and Scope Presentation. NACA RM A57J01, 1958.
11. LAMPKIN, B. A.; AND RANDLE, R. J.: Investigation of a Manual Sextant-Sighting Task in the Ames Midcourse Navigation and Guidance Simulator. NASA TN D–2844, 1965.
12. PRICE, H. E.; SMITH, E. E.; AND BEHAN, R. A.: Utilization of Acceptance Data in a Descriptive Model for Determining Man's Role in a System. NASA CR–95, prepared under contract no. NAS2–1346 by Serendipity Associates, 1964.

Pilot Performance: Research on the Assessment of Complex Human Performance[1]

EARL A. ALLUISI

University of Louisville

If one had the responsibility of monitoring the performance and physiological condition of a vehicle operator such as a pilot or an astronaut, what behavioral information would be necessary in order to represent his current, momentary level of performance? What biological or physiological information would be necessary? How would the two kinds of information be collated? Also, if part of the responsibility was the ordering of a return to base that could be accomplished only in 1 or 2 hours after the order, how could the information be used to predict the operator's performance during that future hour or two?

The fact that there is no set of correct answers to questions such as these demonstrates that little is known concerning the assessment of human performance in operational systems. This problem of performance assessment is probably the most important, the most difficult, and the least-studied problem in human-factors engineering today. It is important because performance assessments must provide the ultimate criteria for the validation of other work.

The final validation of selection and training techniques should depend on the assessment of the performance of men who have been differently selected and trained. The final validation of an improved, human-engineered, man-machine system should depend on such an assessment. The evaluation of the effects of various stresses, the measurement of performance decrements, and the establishment of operational limits and even of optimum operational conditions and procedures all depend on the measurement and assessment of performance.

The task of learning how to assess the performance of an operator in a real system—the pilot in an aircraft, an astronaut in a space vehicle, or even the driver of a truck—has been recognized as a difficult one. The basic problem is that we do not know how to assess an operator's performance on complex, meaningful (i.e., real-world) tasks. Thus, we have no criterion measure(s) around which to design our research, and because of this we are forced to do research on the criterion—to do research to discover how complex performance can be assessed. This may be considered direct research on performance assessment.

Three techniques have been used in direct attempts to solve the problems of performance assessment (ref. 1). They represent a dimension of possible approaches, with simulation techniques (ref. 2) at one end, specific-test techniques (refs. 3 and 4) at the other, and a synthetic-work approach between the two (ref. 5).

The major advantage of full-scale simulation lies in its providing a maximum of face validity; it involves the operator in situations that closely resemble the operational

situations to which generalization is desired. Apart from the questions of economic feasibility that arise from the relatively high cost of simulation studies, there are three important disadvantages: (1) There is the difficulty of assessing the operator's performance in the simulated system. If we could assess this, then we probably would be able to assess it in the operational situation; if we cannot assess it in the operational system, there is little likelihood that we could do so in the simulated system. (2) There is the difficulty of generalization from the simulated system. The more faithful the simulation, the greater the generalization of results to the specific operational system that has been simulated, but the less the generalization to other systems. That is to say, to the extent that the results of the simulation include variances based on specific factors, generalizations can be made to operational systems that also include these specifics, but not to other systems. (3) Measures are taken of the performance of the total system; that is, system-performance measures are used, and operator and hardware performances are confounded.

Specific-test techniques use a test battery that consists of a number of appropriately selected or designed individual tasks. Their major advantage is that the operator's performance on each individual task can be assessed exactly. There are three important disadvantages: (1) These batteries have little or no face validity and, in the absence of criterion measures, this means there is little or no validating information at all. (2) Because there is little or no resemblance between the test situation and the operational situation, there are some additional serious questions concerning the motivation of the subjects or operators. (3) The principal feature of "complex performance" may be its requirement for the time sharing of multiple responsibilities and duties, but such time sharing is missing in the specific test technique.

Between the simulation and specific-test techniques lies a synthetic-work technique that seeks to minimize the disadvantages of the other techniques with as little loss of the advantages as possible. The synthetic-work technique that will be discussed here is based on the measurement of multiple-task performance (MTP) in a synthetic (rather than simulated) work situation under controlled laboratory conditions.

This approach uses a number of time-shared tests that are combined into an MTP battery. It should be possible to generalize the tests or tasks to a wide variety of systems, although their final generality may be dependent on the same sorts of taxonomy, task analysis, and weightings required for generalizing from the factors represented in the specific-test batteries.

The essential requirements for the successful use of an MTP battery in synthetic-work techniques are related to the battery's having relatively high face validity, both in content and in acceptance by operational personnel. The content validity is required to insure the proper generality. The user acceptance is required because it is only on this basis that the inference can be made regarding the operator's viewing the test situation as being essentially like the operational situation, with his behavior in both situations validly placed within the domain of work behavior.

The kinds of functions performed by man in operating the complex systems of today can be categorized into seven major areas as follows:

(1) Watchkeeping, vigilance, and attentive functions, including the monitoring of both static (discrete) and dynamic (continuous) processes

(2) Sensory-perceptual functions, including the discrimination and identification of signals

(3) Memory functions, both short and long term

(4) Communication functions, including the reception and transmission of information

(5) Higher order functions, including information processing, decisionmaking, problem solving, and nonverbal mediation

(6) Perceptual-motor functions

(7) Procedural functions, including such

things as interpersonal coordination, cooperation, and organization.

These, then, are the functions that must be measured with the tasks that constitute the MTP battery if the battery is to have some measure of content validity.

There are several similar MTP batteries in use today. Figure 1 shows the front view of an operator panel used with the battery constructed at the University of Louisville under contractual support of the U.S. Army Medical Research and Development Command. This will be called the BEID (Behavioral Effects of Infectious Diseases) battery.
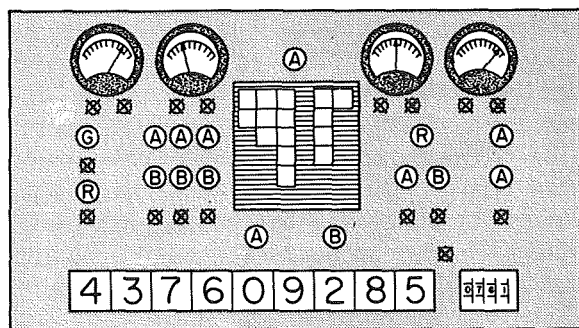
FIGURE 1.—Schematic diagram of the front view of an MTP operator panel. Letters in circles represent indicator lights: A—amber; B—blue; G—green; R—red. The smaller circles with crossed lines represent pushbuttons.

With the BEID battery, behavioral measures are obtained from the operator's performance of six tasks presented with the operator panel. The tasks are displayed on each of five identical panels, one for each member of a five-man crew.

All of the tasks were selected to meet criteria of validity, sensitivity, engineering feasibility, reliability, flexibility, workload variability, trainability, and control-data availability as defined elsewhere (ref. 6). Because each task has been described fully in one or more previous reports (refs. 7–13), detailed descriptions are not repeated here.

Three monitoring tasks are used to measure the operator's performance of watchkeeping, vigilance, and attentive functions.

Blinking lights monitoring.—On the extreme right of the panel (fig. 1) are two vertically arranged amber lights. Under normal conditions, the two lights flash alternately at an overall rate of two flashes per second. The critical signal for which the operator is to be vigilant is an arrest of this alternation in which either the top or the bottom light flashes at twice its usual rate. His latency of response is recorded.

Warning lights monitoring.—On the extreme left of the panel are two warning lights, one green and one red. The operator is required to turn the green light on should it go off, and the red light off should it come on, by depressing a pushbutton located immediately below the light in question. Latencies of these responses are recorded.

Probability monitoring.—Four meters are located along the upper edge of the panel. The pointer on each scale is driven by a random program generator. The pointer positions are normally distributed with a mean of zero (12 o'clock position on the scale) and a known standard deviation. Periodically, the mean of the distribution on one of the four scales is shifted by a specified amount while the variability is unchanged. When the operator detects a shift in the mean, he indicates this by depressing a pushbutton under the meter in question: the left pushbutton if he has detected a bias to the left, or the right pushbutton for a bias to the right. Data recorded are the number of bias signals presented, the number detected correctly, the number of false responses, and the time required to detect each bias correctly.

Three active tasks are used to monitor the operator's performance of memory functions, sensory-perceptual functions, and procedural functions.

Arithmetic computations.—Three three-digit numbers are displayed along the lower central portion of the panel by means of nine one-digit numerical indicators. The operator is required to subtract the third three-digit number from the sum of the first two. He indicates his answer by use of four decade thumb switches immediately to the right of the indicators, and presses a pushbutton just

to the left and slightly above the switches to record his answer. If the answer is correct, a blue indicator light (immediately above the numerical indicators and just to the right of center) is lit for a ½-second interval as the problem is removed and just prior to the presentation of a new problem. Problems are presented at a rate of three per minute. The criterion used in earlier studies was the percentage of responses correctly made by each operator during the performance period. In the later BEID-series studies, two new criteria have been substituted for this: the percentage of problems attempted, and the percentage of problems correctly answered. The arithmetic-computation task measures both short- and long-term memory functions. Of course, it also involves information handling and, to a certain extent, higher order, mediational functions. Thus, it is not a pure measure of memory functioning, but rather it is heavily involved with memory in a manner quite similar to real work. It is also an excellent user of channel capacity and permits realistic loadings of the operator.

Target identification.—A 6 by 6 matrix of square lights is in the center of each operator panel. Contoured figures are generated by lighting selected elements of the matrix. A stored target image is first presented to the operator, followed by two sensed target images, presented in sequence. The stored target always appears as an upright histogram. The sensed targets appear at 0°, 90°, 180°, or 270° from the upright (randomly determined). The operator's task is to indicate, by pressing one of three buttons located below and to the left of the display, whether the first, second, or neither sensed target image is the same as the stored image. Knowledge of results is given to the operator by displaying his response (amber light) and the correct response (blue light) just above the response buttons. Records are made of the total number of responses, and the number of correct responses.

Code-lock solving.—This is a group performance task. It employs three lights (red, amber, and blue) and two pushbuttons (one

of which is a spare) located in the center-right section of the panel, just to the left of the blinking lights. The crew of five men must discover, by trial and error, the correct sequential order for depressing each member's pushbutton in order to complete a trial. Illumination of the red light is the signal that a problem is present and unsolved. The amber light is illuminated on all consoles when any operator depresses his pushbutton, but with no indication as to which operator it was or whether it was just one or more than one who did so. Verbal communication is necessary for this. The red light is extinguished when the correct first operator in the sequence depresses his pushbutton, and it will remain extinguished until an incorrect response is made. When this occurs, the red light is reilluminated, the programing apparatus is reset automatically to the beginning of the sequence, and those group members whose position is known (i.e., those earlier in the sequence) must push their button in the proper sequence to get the group to the point where the error was made. When all five pushbuttons have been depressed in the correct order, the blue light is illuminated as a signal that the problem has been solved.

Following a between-problem pause of 30 seconds, the blue light goes off, the red light comes on, and the crew is presented with a replication of the problem previously solved. This requirement for a second solution has been included to increase the sensitivity of the task to performance decrements. Following the second solution and a between-problem pause of 30 seconds, the blue light goes off, the red light comes on, and the crew is presented with a new sequence or code to solve. Records of the crew's performance are made in terms of the time required for code-lock solutions, the total number of responses made, and the number of errors (or resets of the programing apparatus).

In operating the MTP battery, the concurrent performances of several tasks are required. The intent is to synthesize the several different tasks into a reasonably realistic worklike situation—a situation that requires

an operator to be responsible for more than merely a single function and that permits a variation in workload.

The typical sort of multiple-task performance employed in past studies is shown in table I. The work is divided over a 2-hour performance period so that the operator is responsible for the watchkeeping tasks all the time, but is responsible for the three active tasks only part of the time. Thus, periods of low, high, or intermediate performance demand can be created.

A historical presentation will be used to summarize the results of the research completed with the various versions of the MTP battery. This research was initially aimed at the development of a performance-assessment technique, at the measurement of the effects of working in a volumetrically restrictive environment, and at the determination of optimum work-rest scheduling. More recent work has continued the emphasis on the development of performance-assessment techniques and has begun to measure the behavioral effects of infectious diseases. As spinoff, the data collected have contributed to our knowledge of sustained performance and diurnal rhythms in man.

A program of confinement research was started in 1956 at the Human Factors Research Laboratory of the Lockheed-Georgia Co. under the contractual support of the Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio. The initial concern was with aircrew perform-

ance as affected by confinement of the crew in the anticipated volumetrically restrictive work environment of a nuclear-powered aircraft.

A mockup was designed and constructed to the scale of the nuclear aircraft. However, because of classification and because plans for the vehicle were shelved about the same time that support for the space program was increased, the mockup was changed to accommodate the panels that were then being developed. The total volume was approximately 1100 cubic feet, divided about equally between working and living areas.

The initial plan of study was to use the MTP battery to measure the effects of work-rest schedules on operator performance. The work situation to which generality was desired imposed certain additional constraints on the experimentation; namely, crews rather than individuals would be needed, the work would have to be performed around the clock on a high-alert basis for 5 days or longer, and the work environment would remain volumetrically restrictive (ref. 14). The initial battery was constructed, the crew compartment or mockup completed, and the first data were collected (ref. 15).

The information sought in the first investigation with the MTP battery was concerned with certain technical aspects of its use. For example, the questions asked were concerned with the rates at which operators became proficient on the tasks, the test-retest reliabilities of the measures of per-

TABLE I.—*Basic 2-Hour Task Performance Schedule*

Performance task	15-minute interval in each 2-hour period							
	1	2	3	4	5	6	7	8
Blinking-lights monitoring	X	X	X	X	X	X	X	X
Warning-lights monitoring	X	X	X	X	X	X	X	X
Probability monitoring	X	X	X	X	X	X	X	X
Arithmetic computations		X	X					
Code-lock solving			X	X	X	X		
Target identifications						X	X	
Level of demand...........	Low	Medium	High	Medium	Medium	High	Medium	Low

formance obtained, the interactions among the tasks when performed in various multiple-task combinations, and the interest correlations. The operator panel employed in this first study is shown in figure 2. On the basis of the results obtained, the MTP operator panel was modified, and the scale-position monitoring and tracking tasks were omitted.

The MTP battery had demonstrated impressively high reliability, and it appeared to provide measures of essentially orthogonal functions. It imposed relatively minor training requirements, and was capable of being programed in numerous ways to make possible the study of a broad range of operator workloads (ref. 15). Five tasks were retained. There were three watchkeeping tasks: auditory vigilance, probability monitoring, and warning-lights monitoring; and two active tasks: arithmetic computations and pattern discriminations. The presentation of these tasks was integrated into a 2-hour work period that is shown in figure 3. Again, it should be noted that time-shared performances were required by the scheduling provided in each 2-hour performance period.

The plan of research for a series of 4-day (96-hour) studies is shown in figure 4. The effects of the lengths of the duty and rest periods were investigated first, holding constant the duty-to-rest ratio. Next, the effects of the duty-to-rest ratio were studied, holding the rest period constant. Subsequently, the most efficient work-rest schedule was selected and tested over a 15-day period of performance.

In accordance with this plan, the second experiment sought to measure the effects of the durations of the work and rest periods over a 4-day (96-hour) interval. A unit work-to-rest ratio was used, with work-rest cycles of 2 hours on duty and 2 hours off, 4 hours on and 4 hours off, 6 hours on and 6 hours off, and 8 hours on and 8 hours off. Sixteen male college students served as subjects, with four subjects assigned to each of the four work-rest schedules. The principal data were obtained from the performance measures of the MTP battery, but additional data were obtained from an experimenter's record (logbook) and from a questionnaire administered to the subjects after the test.

The results indicated that, throughout the 96-hour period, the performance scores continued to improve with each of the four work-rest schedules, and no significant difference among the schedules was obtained. The data of the experimenter's log and the questionnaires suggested, however, that the 2- and 4-hour cycles resulted in more favorable adjustments by the subjects than did the 6- and 8-hour schedules (ref. 7).
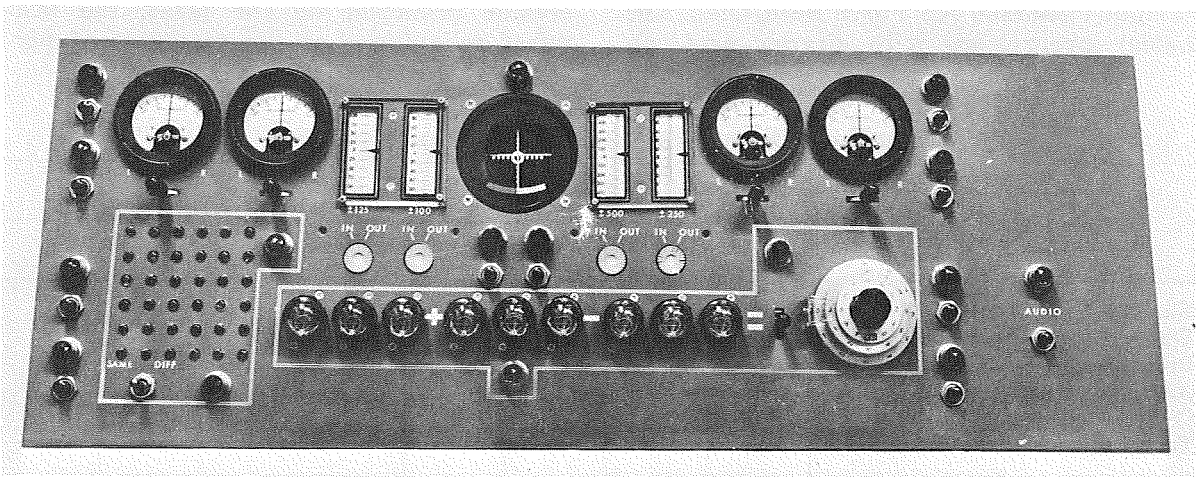
FIGURE 2.—MTP operator panel used in first performance study.
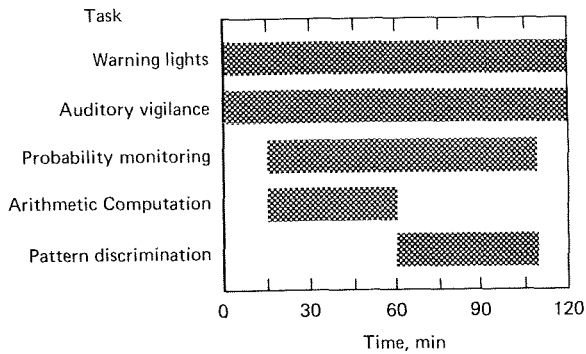
FIGURE 3.—The 2-hour task-performance schedule used with the modified MTP operator panel.
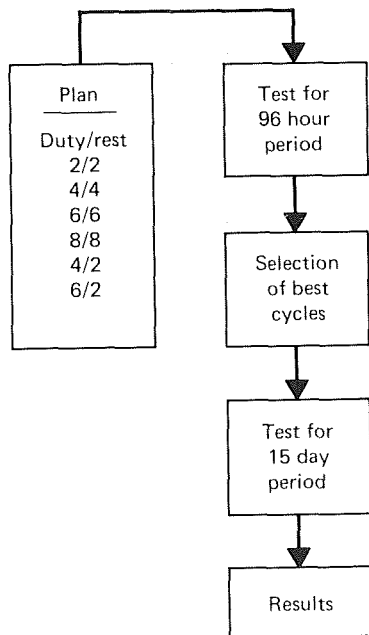
FIGURE 4.—Initial plan for study of work-rest scheduling.

studying the effects on performance of work-to-rest ratios of 2:1 and 3:1.

Twenty male college students served as subjects. They were divided into four groups of five subjects each; two of the groups followed a work-rest schedule of 4 hours on duty and 2 hours off, around the clock for 4 days, whereas the other two groups followed a schedule of 6 hours on and 2 hours off for the same length of time.

The performance data gave clear evidence of diurnal (24-hour) cycling on all measures with both schedules. This is illustrated with the data of the arithmetic-computations task in figure 5; also illustrated is the fact that
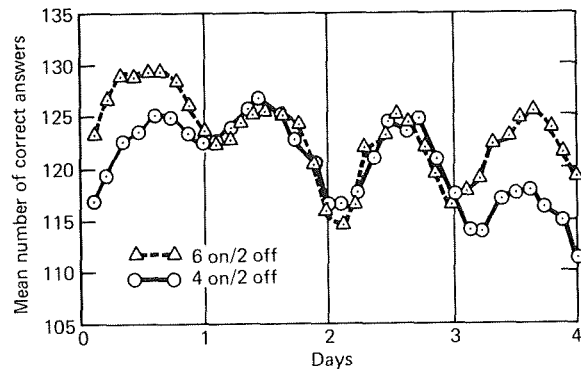
FIGURE 5.—Comparison of the 4–2- and 6–2-hour schedules in terms of number of correct arithmetic computations.

From these data, it appeared that subjects could follow a work-rest schedule that permitted rest (or sleep) periods as brief as 2 or 4 hours in duration. Indeed, the questionnaire data suggested that the subjects preferred the shorter work periods and that they were willing to trade off the length of the rest period to obtain briefer work periods. On the basis of this conclusion, it was decided to use the brief 2-hour rest period in

the differences in the performances obtained with the two schedules did not permit any decision concerning which of the two was the better.

However, there was a clear indication that the subjects preferred the 4–2-hour schedule over the 6–2-hour schedule. In addition, the experimenter's log indicated that the subjects who followed the 4–2-hour schedule averaged at least 5½ hours of sleep per 24-hour period, whereas those who followed the 6–2-hour schedule averaged less than 4 hours of sleep (ref. 8, pp. 29–34).

All of the 4-day studies had one feature in common: they failed to produce differences in performance that could be used to reach meaningful decisions concerning the

efficacy of the various work-rest schedules. This appears to be more a function of the total duration of the studies—that is, the 4-day or 96-hour periods—rather than a function of any lack of effect of the work-rest schedules or lack of sensitivity of the performance measures (ref. 16). Man apparently has the necessary resiliency to meet the demands of quite stressful work-rest schedules, such as the 6–2-hour schedule, over relatively brief 4-day intervals. Presumably, he has performance reserves that he can use to help him over such brief stressful periods. Studies of longer duration appear to be necessary to demonstrate work-rest schedule effects on performance.

The results of the 4-day studies did suggest, however, that a work-rest cycle of 4 hours on duty and 2 hours off might provide a highly efficient schedule for operators who had to perform the kinds of tasks included in the MTP battery. The results did not indicate whether the operators could maintain acceptable performance over prolonged periods of, say, 15 days. This was measured in the first of several long-duration studies with the MTP battery (ref. 8).

Two crews of operational USAF personnel each followed a 4–2-hour work-rest schedule for 15 days. There were five subjects in one crew and six in the other. Both physiological and performance data indicated diurnal cycling (24-hour phase) throughout the 15 days; this is illustrated in figure 6, again with the data of the arithmetic-computations task.

In general, the performance cycles lagged about 2 hours behind the physiological cycles. Also, there appeared to be slight shifts in the cycles throughout the study—shifts that could be interpreted as slightly lengthened cycles of greater than 24-hour periodicity. These shifts, and the lag, are illustrated in figure 7.

This apparent shift is tentatively interpreted as an indication of fatigue, or a work-rest schedule stress; that is, as a result of the accumulated fatigue produced by the demands of the schedule, the subjects were reaching their physiological and perform-
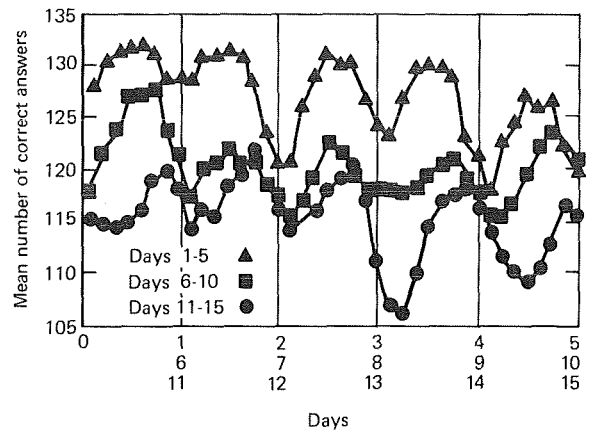
FIGURE 6.—Mean number of correct arithmetic computations on each of 15 days.

ance peaks slightly later each day. This hypothesis may provide some interesting measures of fatigue and work-stress effects, if it is validated in other long-duration studies.

In general, the data did support the hypothesis that the 4–2-hour work-rest schedule could be used. Specifically, it was concluded that, with some selection, highly motivated crews could maintain acceptable performance levels while following a 4–2-hour work-rest schedule for a period of 2 weeks, and possibly for longer durations. The conclusion was based principally on the fact that 2 of the 11 subjects were able to maintain high performance levels throughout the 15 days. In addition, the majority of subjects indicated during posttest interviews that they could have continued the test for at least an additional 15 days if it were necessary and important for them to have done so.

The crew-performance code-lock task was subsequently developed, and the pattern-discrimination task was replaced with target identifications (refs. 9 and 10). The resultant panel is shown in figure 8; the 2-hour performance periods employed with this panel were the same as those currently used with the BEID battery (as was shown in table I), with the exception, of course, that blinking-lights monitoring has replaced auditory vigilance in the final version.
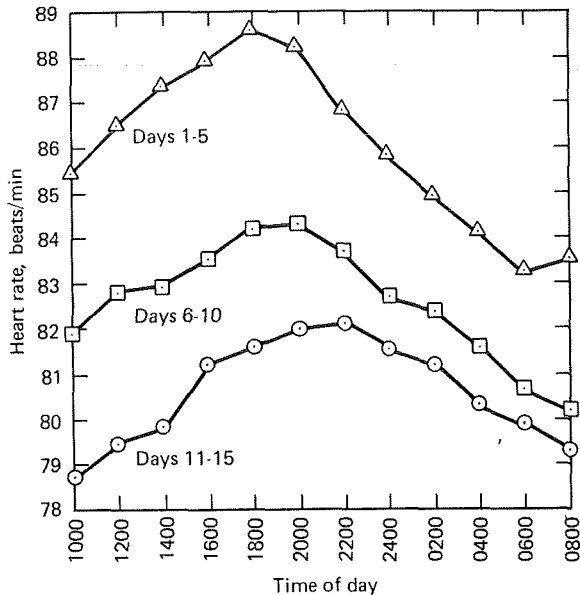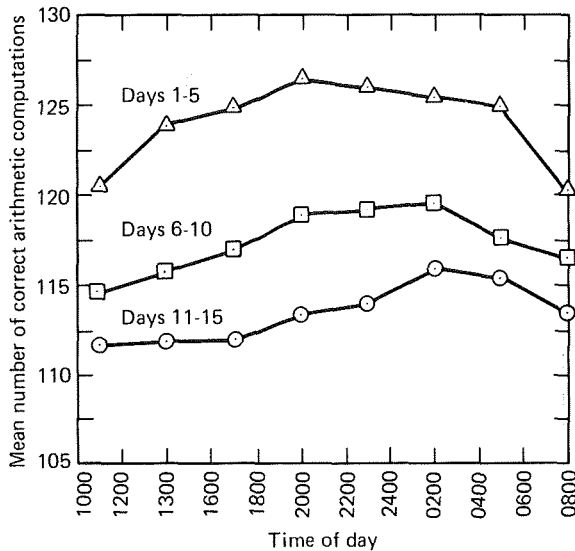
Figure 7.—Shifts in performance cycle. (*a*) Within-day changes in heart-rate level. (*b*) Within-day changes in level of correct arithmetic computations.

FIGURE 8.—MTP operator panel as modified by addition of crew-performance tasks (target identification and code-lock solving).

Besides the newly added crew-performance code-lock task, the target-identification task was provided with a secondary display at the crew commander's position so that it also reflected certain aspects of crew, rather than simply individual performance. Then,

with the addition of these crew-performance measures to the MTP battery, the 15-day study of the effects of the 4–2-hour work-rest schedule was replicated. The subjects were six highly motivated Air Force Academy cadets—probably the most highly motivated subjects this experimenter has ever encountered! The subjects were asked to do whatever they could do to prevent the expected diurnal cycling in performance by expending extra effort during those work periods that seem to be hard on you—usually those in the early morning hours between, say, 2:00 and 6:00 a.m.

The physiological data (self-determined axillary temperatures and pulse rates) gave clear indications of a diurnal rhythm (24-hour period), whereas the performance data generally indicated no diurnal cycle. This is illustrated by the solid curve in figure 9, the axillary-temperature data of this group of subjects who were referred to with the code name "HOPE–II." The broken curve shows the skin-temperature data of the operational crews run in the previous 15-day study, referred to as "OPN–360."

The general absence of diurnal cycling in the performance data of HOPE–II, especially relative to that which was evidenced in the data of OPN–360, can be seen in figure 10 with the data of arithmetic computations (percentage of correct responses, when the task was not time shared with the code-lock task). Also shown are the data of a group of 10 subjects, HOPE–III, who followed a work-rest schedule of 4 hours on duty and 4 hours off for a 30-day period. It can be seen that
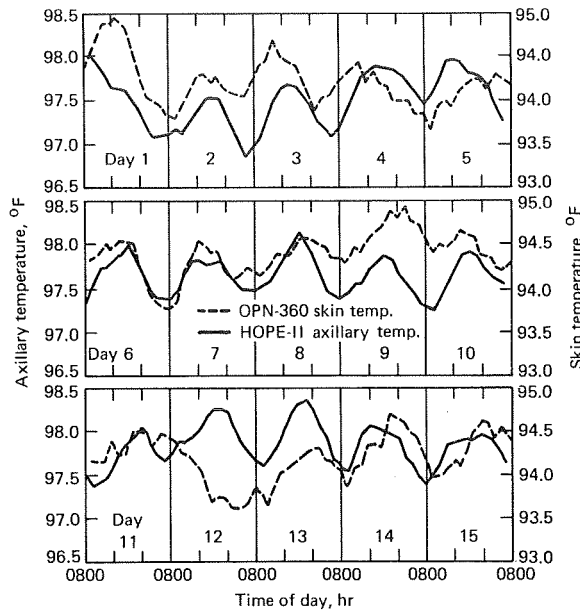
FIGURE 9.—Mean temperature data of OPN–360 and HOPE–II.

FIGURE 10.—Mean percentages of correct arithmetic computations without concurrent code-lock solving: OPN–360, HOPE–II, and HOPE–III.

the subjects of HOPE–III, like those of HOPE–II, produced performance data that generally failed to indicate diurnal cycling.

There was one important exception to this latter result in the data of both HOPE–II and HOPE–III; this exception will be discussed, but first certain conditions of the 30-day HOPE–III study need to be explained further.

The 10 subjects in the 30-day study were USAF pilots. They were divided into two five-man crews that followed the 4–4-hour work-rest schedule for the 30-day duration of the study. The subjects were led to believe, however, that the confinement for the study would extend for 40 days, and, because they did not learn otherwise until the study ended on the 30th day, these data safely can be assumed to show no end effects.

The physiological data give clear evidence of diurnal cycling; this is illustrated in figure 11, where the self-determined axillary temperatures are shown for the 30-day period. Here, as in figures 9 and 10, rolling means were used to minimize the effects of variations attributable to differences among individuals and work activities. Thus, each

point has represented in it all 10 subjects and an equal number of subjects who have just arisen from sleep, who are in the midst of a 4-hour work period, and who have just completed a 4-hour work period.

Several conclusions appear to be supported by the data of figure 11. First, the diurnal cycling with 24-hour periodicity in axillary temperature is clearly evident. Second, the drifting or lagging of the diurnal cycle— suggestive of a cycle that is slightly longer than 24 hours—is also evidenced; this is shown in the peaks of the broken line's being slightly displaced to the right of the peaks of the solid line. Third, it can be seen that the diurnal cycling in temperature apparently continued without much abatement for the first 20 or 25 days of the study. Only during the last 5 or 10 days does the diurnal variation appear to be somewhat flattened. This suggests that physiological adaptation to atypical work-rest schedules will take at least 20 days, and perhaps as long as 25 or
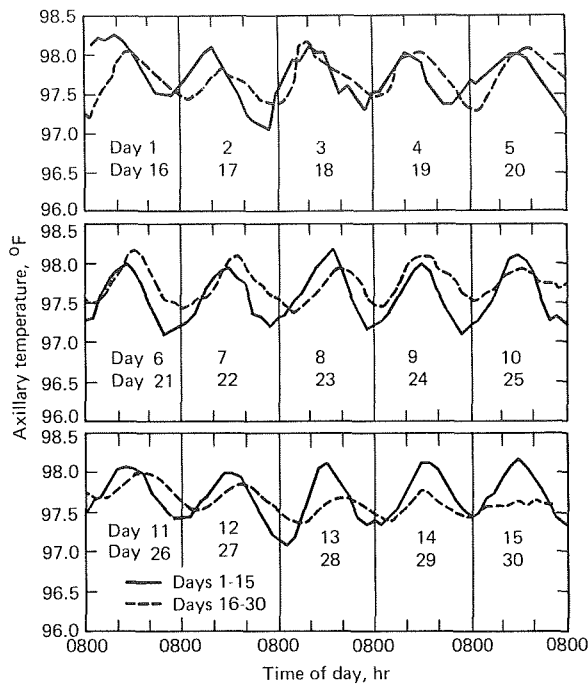
FIGURE 11.—Mean temperature data of HOPE–III: solid line indicates data of days 1–15, and broken line days 16–30.

30 days on the average—even where the environment is controlled and the variations due to activity are removed by means such as the rolling-mean techniques used here. There is additional support for this conclusion in the literature (ref. 16).

The HOPE–III subjects, like those in the 4–2-hour HOPE–II study, had been shown the data of the first 15-day experiment (OPN–360) in which diurnal variations in performance had been noted. They, too, were instructed to expend extra effort when necessary to preclude such cycling effects. Although these subjects did exhibit significant performance cycling on some of the tasks, the magnitude of the effect was not great; and, as was previously shown in the arithmetic-computations data of figure 10, the lowest levels in the HOPE–III cycles still represented substantially better performance than that generally exhibited by the OPN–360 subjects on the 4–2-hour schedule.

As indicated earlier, there was one important exception. This is shown in figure 12

with the arithmetic-computations data that were collected while the operators were concurrently performing the code-lock task. The data given are those of the second, less-demanding, 4–2-hour study (HOPE–II) and those of the even easier 4–4-hour study (HOPE–III). The data indicate a condition of performance stress for both groups during the first several days of experimentation, apparently while they were still learning to time-share the tasks.

Performance stress may be said to exist when, on the introduction of an additional task, performance on all tasks (including the newly introduced one) falls below the levels attained without the additional task. This was the case during the early days of performance when the code-lock task was added to the demands of the arithmetic-computations and watchkeeping tasks. It is apparent, for example, from the bottom panel in figure 12, where the data of the third and sixth 5-day periods of HOPE–III are shown along
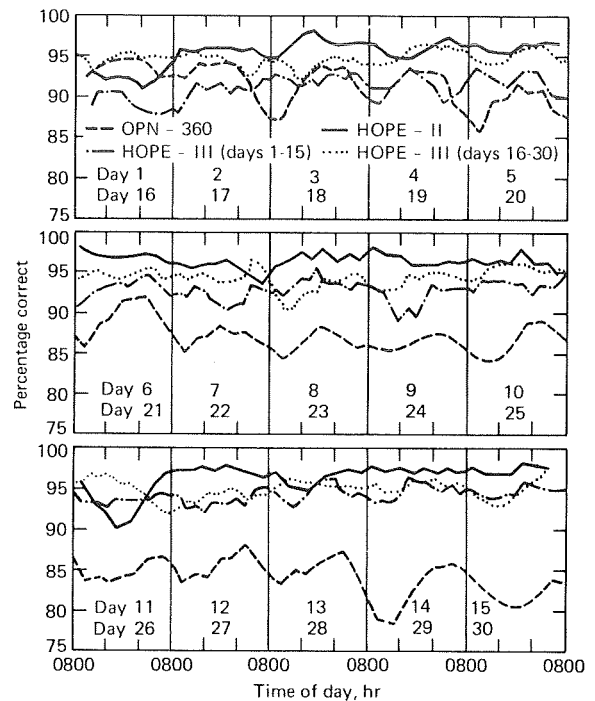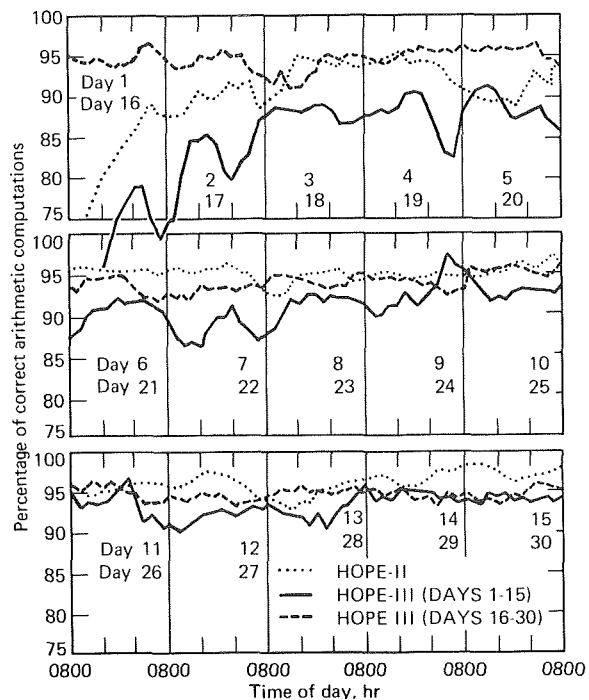

FIGURE 12.—Mean percentages of correct arithmetic computations with concurrent code-lock solving: HOPE–II and HOPE–III.

with the third period of HOPE–II, that the subjects did eventually learn to time-share these tasks. Also, because these levels of performance are essentially identical with those obtained without simultaneous presentation of the code-lock task (fig. 10), it can be concluded that the condition of performance stress no longer existed.

Rather, it was during the first several days that the performance stress was evidenced. During those days, the diurnal cycling of performance was clearly indicated in the data. This is interpreted to mean that even when the subjects are generally able to overcome diurnal-cycling effects in their performances (the data of fig. 10, arithmetic computations without simultaneous presentation of the code-lock task, showed no diurnal-cycling effects even during the early days of practice), they are able to do so only within limits. A physiologically determined diurnal rhythm is present and underlies all performance; information and motivation can be used to overcome the tendency for performance to exhibit the same rhythm, but only to a point. If the subjects are overloaded —if they have more than they can do, as in the performance-stress condition—the diurnal-cycling effects are likely to reappear in the performance data.

The results of the 30-day study otherwise indicated that the 4–4-hour work-rest schedule was less demanding than the 4–2-hour schedule (ref. 10). It was concluded that, whereas with proper control of selection and motivational factors, crews can work effectively for at least 2 weeks (and probably longer) using a schedule of 4 hours on duty and 2 hours off, crews can work even more effectively for periods of at least 1 month (and quite probably for 2 or 3 months) using a schedule of 4 hours on duty and 4 hours off. Also, the latter schedule would apparently require less-demanding controls of the selection and motivational factors.

These conclusions were further supported by the results of four 12-day studies of the combined effects of sleep loss and the two work-rest schedules (ref. 11). This is illustrated in figure 13 in terms of the percentage
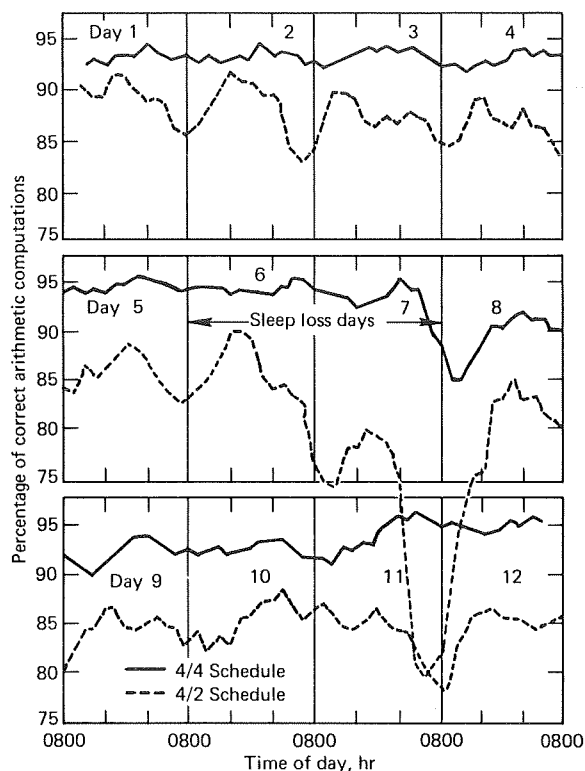
FIGURE 13.—Mean percentages of correct arithmetic computations without concurrent code-lock solving: 4–4- and 4–2-hour work-rest schedules with sleep loss.

of correct responses in the arithmetic-computations task when it was presented without simultaneous presentation of the code-lock task. The solid curve represents the data of the two groups (20 subjects) who followed the 4–4-hour work-rest schedule, and the broken curve represents the two groups (12 subjects) who followed the 4–2-hour schedule.

Performance was generally inferior on the 4–2-hour schedule as compared with the 4–4-hour schedule, and the stress of sleep loss (40 and 44 hours of wakefulness with the two schedules, respectively) resulted in greater performance decrements for subjects on the 4–2-hour schedule than for those on the 4–4-hour schedule. Diurnal cycling in the performance measures of the 4–4-hour subjects was generally not apparent except during the period of sleep-loss stress.

The performance of the 4–2-hour subjects also showed increased diurnal cycling during the sleep-loss period. Thus, it again appears that the extent to which motivation can be used to overcome the physiological rhythm is limited; operators, no matter how highly motivated, cannot achieve the impossible!

Several control studies have also been conducted, although direct reference has not been made to them. In one (ref. 8, pp. 35–39), a group of six college students was tested 4 hours a day, 5 days a week, for 6 weeks (120 hours of performance). When their performance was compared with that of the first 15-day 4–2-hour study (OPN–360), it was generally found that the controls continued to improve and were superior to the experimentals on most tasks (specifically, on arithmetic computations, warning-lights, and auditory-vigilance tasks).

A second control study (ref. 17) was concerned with an evaluation of performances on the three watchkeeping tasks (auditory vigilance, probability, and warning-lights monitoring) when they occurred with and without simultaneous presentation of the three active tasks (arithmetic computations, target identifications, and code-lock solving). After familiarization and preliminary training given over a 4-day period, two groups of five college students were tested for 4 hours per day on 6 successive days. It was concluded that concurrent presentation of the active tasks has a detrimental effect on the operator's performance of his watchkeeping duties. The effects are similar to those of performance stresses, in that removal of the concurrently presented active tasks invariably resulted in the recovery of the watchkeeping performances back to the previously attained levels.

The most recent control study was conducted at the University of Louisville, with the UL–Army or BEID version of the MTP battery. In this study, 10 Air Force ROTC cadets followed a work-rest schedule of 4 hours on duty, 4 hours off, 4 hours on, and 12 hours off, for 11 consecutive days. These subjects were not restricted in their activi-

ties except, of course, during the 8 hours of work per day; during the remaining 16 hours, they were free to conduct their normal activities (school was not in session).

The dotted curve in figure 14 presents the control group's percentage of correct responses in the arithmetic-computations task, when this task was worked without concurrent presentation of the code-lock task. The broken line presents the comparable data of the first 12 days of performance of the 10 pilots in the 30-day study of the 4–4-hour work-rest schedule (HOPE–III).

It is apparent from the data in figure 14, as it was in the other data obtained, that the two groups performed at essentially identical levels. From this, it can be concluded that the 4–4-hour schedule under the conditions of the 30-day study (i.e., with controlled environmental conditions) produced performances that were as good as those obtained with a less demanding 4–4–4–12-hour
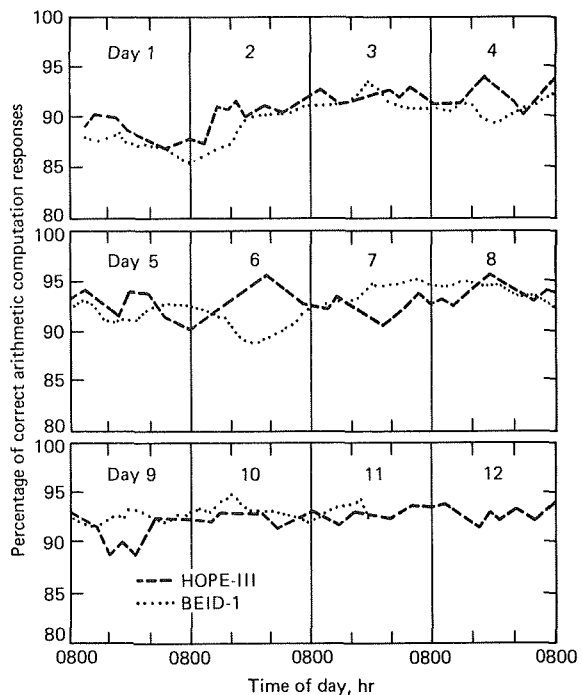
FIGURE 14.—Mean percentages of correct responses to arithmetic-computations problems without concurrent code-lock solving: BEID–1 and first 12 days of HOPE–III's performance.

schedule under the more nearly normal working conditions in which the operators were free during their off-duty hours to do what they wished (ref. 12).

The men who operate our modern, complex, man-machine systems are subject to illnesses; they are subject to infection as long as they live. A large part of the biomedical literature is devoted to infectious diseases, and a great deal of medical practice is devoted to the prevention of infection and the treatment and cure of infected individuals.

Relatively little of the psychological literature, however, has been concerned with the effects of infectious diseases on behavior. If only little is known of man's behavioral reactions to infection, essentially nothing is known regarding the quantitative effects of infectious diseases on his performance or work (ref. 18). There is little doubt that this dearth of knowledge concerning the behavioral effects of infectious diseases is based in great part on the basic need for a suitable performance-assessment methodology.

It is not unnatural, therefore, that the present line of MTP research on the general question of performance assessment has turned specifically to the study of the effects of illness on performance. Hopefully, the research will lead us closer to both goals.

The first experimental study was conducted at the U.S. Army Medical Unit, Walter Reed Army Medical Center, Fort Detrick, Md., in January 1966 with 10 volunteer subjects, 8 of whom were infected with respiratory *Pasteurella tularensis* (commonly called tularemia or rabbit fever) and two of whom served as uninfected controls. The subjects worked on a schedule of 4 hours on duty, 4 hours off, 4 hours on, and 12 hours off, during each of 12 successive days. Exposure occurred on the morning of the fifth day of testing. As indicated by the rectal temperatures shown in figure 15, the experimental subjects were febrile by the 8th day, remained so during the 9th day when treatment was started, went through normal on the 10th day, and were slightly subnormal in temperature on the last 2 days of testing.
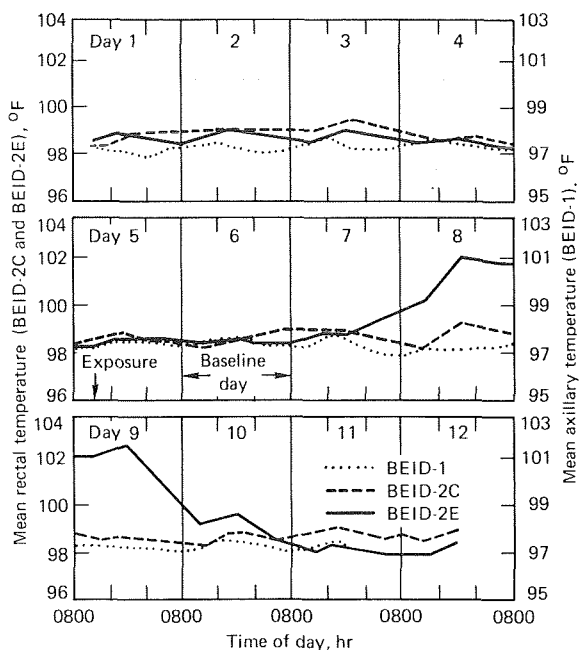
FIGURE 15.—Mean temperature data of BEID–1 and BEID–2. The axillary temperatures of BEID–1 are read from the ordinate on the right; rectal temperatures of BEID–2C (controls) and BEID–2E (experimentals) are read from the ordinate on the left.

The broken curve in figure 15 presents the rectal temperatures of the 2 control subjects; the dotted curve presents the axillary temperatures of 10 subjects in the University of Louisville control group that was discussed previously.

Respiratory tularemia is a febrile disease based on infection with the *Pasteurella tularensis* bacterium. It is characterized by severe headache, photophobia, nausea, myalgia, and depression. All eight of the experimental subjects became symptomatic on either the seventh or eighth day of testing, and chemotherapy was begun on each during either the eighth or ninth day. Both of the double-blind controls remained asymptomatic throughout the period of testing.

Decrements in the performances of the experimental subjects were measured during the period of illness with each of 13 scores that are based on the 6 tasks in the UL-Army MTP battery. In addition, a general index

of performance was derived in order to represent in a composite score the general performance of the subjects on all tasks. This index of general performance is the mean percentage of baseline performance, where each subject's performance with each of the 13 measures during the sixth day of testing was taken as the baseline and set at 100 percent. The results obtained with this index of general performance are shown in figure 16.
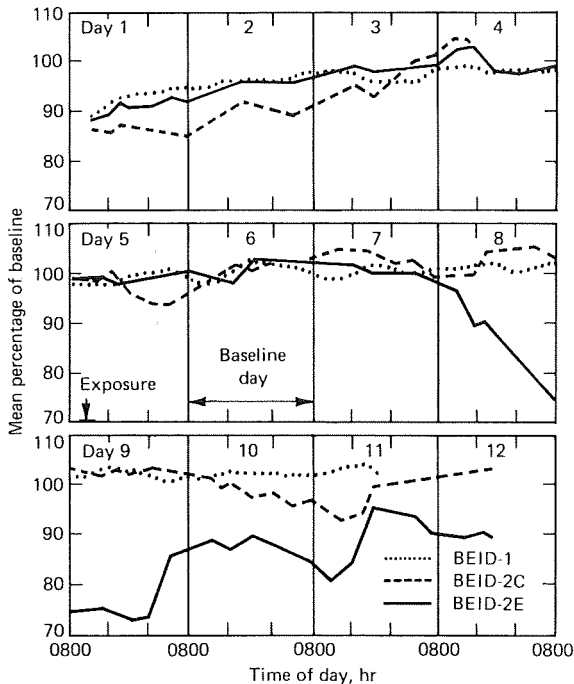
FIGURE 16.—Mean percentages of baseline (an index of general performance) : BEID–1 and BEID–2.

As is indicated in figure 16, average efficiency fell about 25 percent during the period of illness. Recovery was incomplete 3 days after treatment had begun, with performance averaging 10 to 15 percent below that of control subjects (ref. 12).

In terms of the individual tasks, illness-related decrements in performance were evidenced more clearly in the active tasks relative to the watchkeeping tasks, but the

recovery to baseline performance following treatment was less nearly complete in the watchkeeping tasks than in the active tasks.

During the period of illness, there was on the average a 6 percent drop in performance efficiency with each 1° F rise in rectal temperature. However, individual differences among the subjects were very great; they varied from essentially no decrement in performance to a decrement of about 20 percent per degree in the case of one subject. Additional research will be required to identify the psychological and biomedical correlates of such performance decrements. It is hoped that such research will lead not only to an understanding of the differences among individuals in their behavioral reactions to illness but also to continued progress toward suitable solutions to the problems of performance assessment.

An attempt has been made here to summarize the philosophy, techniques, and data of a program of research on performance assessment, which began in 1956. The methodology developed has employed a synthetic-work situation in which it is possible to assess the performances of operators (subjects) who are required to work at the time-shared tasks presented with an MTP battery. The tasks themselves were selected to measure certain behavioral functions that man is called upon to perform in a variety of today's complex man-machine systems. Specific research studies have dealt with confinement in a volumetrically restrictive environment, sustained performance, work-rest scheduling, the behavioral effects of infectious diseases, and diurnal rhythms in man.

The general conclusions reached are the following:

(1) Crews consisting of as many as 10 men can be confined in a space as small as 1100 cubic feet for as long as 30 days or more without observable detrimental effects.

(2) Men apparently can follow a work-rest schedule of 4 hours on duty and 4 hours off for very long periods without detriment to their performances.

(3) For shorter periods of 2 or possibly 4 weeks, selected men can follow a more de-

manding 4–2-hour work-rest schedule with reasonable maintenance of performance efficiency.

(4) In following the more demanding schedule, man uses up his performance reserve and so is less able to meet the demands of emergency conditions, such as those imposed by sleep loss.

(5) The diurnal rhythm, which is evidenced in physiological measures, may also be evidenced in the performance, depending also on the total workload.

(6) Even when motivation is sufficiently high, the diurnal cycling of performance may be demonstrated when the operator is overloaded or stressed.

(7) The average performance efficiency of a crew of men will drop about 25 percent during a period of illness with a febrile disease such as tularemia.

(8) During such an illness, the average drop in performance efficiency is about 6 percent per 1° F rise in rectal temperature, but individual differences will be very great and may be expected to range from no decrement to about 20 percent per degree.

(9) The synthetic-work methodology and the MTP battery that were developed and employed in these studies yielded measures that are sensitive to the manipulation of both obvious and subtle experimental variables, and they provided results on the basis of which inferences could be drawn and conclusions reached, such as those listed above. In short, some progress is being made on the assessment of complex human performance.

DISCUSSION

STANLEY DEUTSCH: Why did you switch from audio vigilance to visual vigilance?

EARL A. ALLUISI: Specifically because of the environment in which I had to run the infectious disease studies; namely, the hospital environment where they wanted to keep the auditory channels free.

RICHARD C. ATKINSON: You mentioned earlier that performance improved over days. . . .

ALLUISI: I mentioned that I could not distinguish between the two schedules with the performance data. They both dropped off, but they dropped off to the same degree and I could not say that the 4–2-hour schedule was better than the 6–2-hour schedule on the basis of the performance data obtained in 4 days.

There were some differences, as I indicated. When the subjects came out of the 4-day studies, the 4–2-hour-schedule subjects came out essentially the same way they went in; that is, they were relatively friendly toward each other and toward the experimenters. We asked at the end of the study during the debriefing, sort of operationally, "How did you like the study?" We asked it this way (these were college students): "We may need to replicate this study next month. If we do, it would be to our advantage to use people who have been trained. Would you like to serve again? If so, just check yes." All of the subjects in the 4–2-hour schedule said, "Well, sure, if you pay us." (We had paid them for 24 hours' work a day since on each day they worked and were confined for 24 hours.) In 4 days they made nearly as much money as they would normally get in a quarter's part-time work.

We asked the 6–2-hour-schedule people this. They informally answered, "Hell, no." When they came out, they were pretty stressed, and the clinical psychol-

ogist who looked at their behavior described it as stress behavior. They were easily angered, short tempered, not getting along with each other, and not getting along with the experimenters. One subject nearly destroyed a panel because he missed a problem—just hit it, banged on it. The subjects were beginning to show the typical effects of a sleep-loss stress.

We ran the study before we completed the literature review. If we had done the literature review first, I would have been able to predict that we would not obtain definitive results in a 4-day study. The literature indicates that in order to demonstrate a work-rest scheduling effect, you must go at least 1 week, and preferably 2 weeks, before you can expect an effect.

JOSEPH MARKOWITZ: How much training was there before they entered into the 4 days?

ALLUISI: We used 2 days of what I would rather call familiarization training. We trained them to use the battery and then put them to work right away. They had one-half hour of training on each task by itself, and then another half hour on each in combination with other tasks before we began data collection. In this sense it was familiarization training.

There is a reason for our having begun our studies immediately following the familiarization training: remember, these were work-rest scheduling studies and all the schedules were difficult. In order to make sure that our subjects (adult human beings) would follow the assigned schedule, we did not leave them free to follow any other schedule. So we did not try to adapt them to the schedule before they went into the experiment. In the pretraining or familiarization period, the subjects worked during the day, but dur-

ing the evening they were free to go to Atlanta and do what people do when they go to Atlanta. That is not the way to break into a new work-rest schedule. So, in order to have the subjects follow the schedule, we put them in the mockup and started the study. In all these cases, however, the first 5 or 6 days of full data collection in the mockup are days of training in the usual sense: they are trained up to a baseline of performance, and in the cases of work-rest studies, these are days of adaptation to the new work-rest schedule.

MARKOWITZ: Would you say they were well trained by the end of the first day?

ALLUISI: No; but let me explain. Each task was selected in part because it did not show much of a learning effect. As we told our subjects, "Believe or not, you can add now as well as you will be able to add at the end of this study. All the problems you are going to complete will not increase your skill at adding and subtracting. What you will learn is how to time-share mental arithmetic with other activities." Thus, the subjects learned the work; they learned the job; they learned to share the arithmetic with the other things demanded of them. They did learn to time-share the tasks, and in general they were on asymptotic levels of performance by the end of the fifth day. Of course, the previous 4 days also represented to some extent an adaptation period to the work-rest schedule, so the learning of time sharing was confounded with the adaptation to the schedule.

ATKINSON: Those cycling functions are really quite impressive. If you had started them at 12 noon as opposed to 8 a.m. in the beginning of the experiment, do you think the function would depend on the starting time or on the actual time of day; that is, it increases from 8 o'clock up and then drops down again, and so forth.

DEUTSCH: I have a corollary comment to that. You indicated that at noon and late afternoon their performance would increase. I notice on your chart that as the period of confinement increases, the increased performance occurs later and later in the day.

JOHN W. SENDERS: The heart rate topped 88 beats/min. Why was it at such a high level? Eighty-eight is extraordinarily high for the small amount of physical activity.

ALLUISI: I have a hypothesis! The next group of subjects that we ran were Air Force Academy cadets who were flown directly from Denver to Atlanta. In their case, the heart-rate level changes in exactly the opposite direction: it starts low and ends up high. In their case, we suspect this change to have been an adaptation to the change in altitude.

There is one report that I have run across by a flight surgeon with SAC who indicates that during the period in which this study was run (1959), SAC crews tended to go into a state of physiological alert about 2 hours before a mission and to remain that way until about 4 hours after mission completion. By physiological alert, I mean elevated pulse rate, and so on.

The two SAC crews in our study had been flying average 4-day missions. We believe they went into our study with an elevated pulse rate in something like a physiological conditioning to flying a mission. Then, after the first 4 days, the pulse rate slowly returned to normal because this mission did not stop, it just kept on going.

SENDERS: It is really surprising, because I presume their physical condition was first class, if they were regular SAC crews. And I would have imagined that a resting heart rate of something in the order of 66–68....

STEVEN E. BELSLEY: They are not resting.

SENDERS: They are not doing physical labor of any great amount. They are sitting down while they are pushing the buttons. In your heart-rate monitoring, did you also get a measure of heart-rate variability over the short term? I am thinking of the work of Carlsbeck at the Dutch Research Institution at Amsterdam, on what he called sinus arrythmia and the claimed relationship between workload and heart-rate variability as opposed to a variation in mean rate itself.

ALLUISI: Yes. As I recall, heart-rate fluctuations were obtained. I do not recall anything extraordinary about it.

REFERENCES

1. CHILES, W. D., ED.: Conference Proceedings: Assessment of Complex Operator Performance. Human Factors, vol. 9, 1967, pp. 325–392.
2. GRODSKY, M. A.: The Use of Full-Scale Mission Simulation for the Assessment of Complex Operator Performance. Human Factors, vol. 9, 1967, pp. 341–348.
3. FLEISHMAN, E. A.: Performance Assessment Based on an Empirically Derived Task Taxonomy. Human Factors, vol. 9, 1967, pp. 349–366.
4. PARKER, J. R., JR.: The Identification of Performance Dimensions through Factor Analysis. Human Factors, vol. 9, 1967, pp. 367–373.
5. ALLUISI, E. A.: Methodology in the Use of Synthetic Tasks to Assess Complex Performance. Human Factors, vol. 9, 1967, pp. 375–384.
6. ALLUISI, E. A.; AND FULKERSON, S. C.: Behavioral Effects of Infectious Diseases: Semiannual Progress Report, 1 March 1964–31 August 1964. Rept. no. PR–64–2, Univ. Louisville Performance Res. Lab., 1964.
7. ADAMS, O. S.; AND CHILES, W. D.: Human Performance as a Function of the Work-Rest

Cycle. Tech. Rept. no. 60–248, USAF Wright Air Develop. Ctr., 1960.

8. ADAMS, O. S.; AND CHILES, W. D.: Human Performance as a Function of the Work-Rest Ratio During Prolonged Confinement. Tech. Rept. no. 61–720, USAF ASD, 1961.

9. ALLUISI, E. A.; HALL, T. J.; AND CHILES, W. D.: Group Performance During Four-Hour Periods of Confinement. Tech. Doc. Rept. no. 62–70, USAF AMRL, 1962.

10. ALLUISI, E. A.; CHILES, W. D.; HALL, T. J.; AND HAWKES, G. R.: Human Group Performance During Confinement. Tech. Doc. Rept. no. 63–87, USAF AMRL, 1963.

11. ALLUISI, E. A.; CHILES, W. D.; AND HALL, T. J.: Combined Effects of Sleep Loss and Demanding Work-Rest Schedules on Crew Performance. Tech. Doc. Rept. no. 64–63, USAF AMRL, 1964.

12. ALLUISI, E. A.; THURMOND, J. B.; AND COATES, G. D.: Behavioral Effects of Infectious Diseases: Respiratory *Pasteurella tularensis* in Man. Rept. no. ITR–67–6, Performance Res. Lab., Univ. of Louisville, 1967.

13. CHILES, W. D.; ALLUISI, E. A.; AND ADAMS, O. S.: Work Schedules and Performance During Confinement. Human Factors, vol. 10, 1968, pp. 143–195.

14. ADAMS, O. S.: Aircrew Fatigue Problems During Extended Endurance Flight. Phase I: Planning. Tech. Rept. no. 57–510, USAF Wright Air Develop. Ctr., 1958.

15. ADAMS, O. S.; LEVINE, R. B.; AND CHILES, W. D.: Research to Investigate Factors Affecting Multiple-Task Psychomotor Performance. Tech. Rept. no. 59–120, USAF Wright Air Develop. Ctr., 1959.

16. RAY, J. T.; MARTIN, O. E., JR.; AND ALLUISI, E. A.: Human Performance as a Function of the Work-Rest Cycle. NAS–NRC Publ. No. 882, 1960.

17. HALL, T. J.; PASSEY, G. E.; AND MEIGHAN, T. W.: Performance of Vigilance and Monitoring Tasks as a Function of Workload. Tech. Rept. no. 65–22, USAF AMRL, 1965.

18. WARM, J. S.; AND ALLUISI, E. A.: Behavioral Reactions to Infection: Review of the Psychological Literature. Perceptual and Motor Skills, vol. 24, 1967, pp. 755–783, Monogr. Suppl. 4–V24.

Comments[1]

Douwe B. Yntema
Lincoln Laboratory

Douwe B. Yntema: I would like to comment on Professor Alluisi's talk first, then come back to Mr. Rathert's, and then go off on a tangent of my own.

Mr. Rathert posed us a rather difficult question when he set forth a classification of decisions and then said he did not know whether that was the way psychologists would classify them. I feel that I must try to respond to that challenge. It is a fairly difficult task, so let me postpone it a little and come back to it.

In a way, Professor Alluisi has left me with a much easier job: he has already summarized a long series of experiments and set them in their context. That means a discussant does not have much work to do. There are just two points I might venture to add. In a sense they are related to each other. The first is the great difficulty there has been, historically, in doing research of this kind. The second is a comment on the way that results of this sort might feed into full-scale, realistic simulations of the sort he mentioned.

Historically, this has been a very tough field in which to work. I am not referring to the obvious difficulties—the almost heroic level of devotion demanded of the subject, or the almost equally heroic dedication of an experimenter who sets up such an operation, keeps it running for weeks, and then is faced with a mountain of data when he is through. Rather, I am referring to the fact that as recently as World War II, people did not know how to do experiments in this area. There were lots of people working manfully to show any kind of effect of stressful conditions on the human higher mental processes— an effect of excessive heat, or excessive noise, or sleep loss, on the learning of lists of nonsense syllables, or on mental arithmetic, or on anything else. They were not very successful.

The breakthrough came at Cambridge when Mackworth introduced the vigilance tasks, which required a subject to pay unremitting attention to something or other for hours on end. These tasks did indeed show a decrement during the course of a long watch, and they have since shown decrements with other sorts of variables. More recently, time-sharing tasks that require the subject to divide his attention among two or more subtasks (thus presumably draining off some of his capacity to perform any one of the subtasks) have provided a much more sensitive weapon for investigating the effects of stressful conditions on human higher performance. I think it is not exaggerating to say that 25 years ago it would not have been possible to obtain the sort of results we have seen today.

This is still a very rough area in which to do research. I think the rabbit-fever experiment points that out. Here is a man in a physiological state that would give a flight surgeon conniptions at the thought of his

31

even walking past an airplane: he has a splitting headache, he cannot look at the light, he is nauseated, he aches all over, and he generally feels depressed. But we saw that even with the sensitive measuring tools now available, there was only a 25- or 30-percent decrement in performance. Clearly, the lesson is that further efforts on developing tasks that would prove even more sensitive to adverse conditions are going to be useful in making research more possible in this area. Professor Alluisi has indicated that he and people associated with him are in the process of developing further talks, and that is obviously a good thing.

The second point, the relation of this sort of experiment to the more specific, more realistic, full-scale simulations, is this: I suspect that people who work on such simulations will look back on this work as telling them how to do decent experiments, especially in simulating missions of long duration. In setting up such a simulation, they will know some of the things that they must control if they are to get stable and reproducible results, or, to turn it around, some of the things they may deliberately change in order to get big changes in behavior. They will know that they have to look for diurnal variations. They will have some notion of reasonable watchkeeping schedules to investigate in a full-scale simulation. Generally, they will know how to run a good experiment.

In particular, time sharing has suggested the following technique for investigating the level of difficulty of some task: set up a realistic full-scale simulation and then introduce subsidiary, artificial tasks, and require the subjects to share their time between those tasks and their normal duties. This should drain off enough of their performance reserve so that you can see how difficult the normal duties actually are.

I think it is worthwhile mentioning here the parallel with one of the oldest and most successful branches of human-factors engineering; namely, the transmission of speech by telephone. Forty years or more ago when Fletcher was tackling that problem at the

Bell Laboratories, it was, I gather, a real can of worms. People smiled gently and said you could never get any place with a problem like that. In the real, practical situation, there are just too many factors that affect the satisfactory transmission of speech. You could not begin to sort them out, or show how they all interact with each other.

Yet, now a second-year graduate student who sat down and read perhaps five or six well-chosen references could do a perfectly competent job of setting up and running off a comparison of, say, the effectiveness of two telephone intercom systems. The background knowledge of what matters in the field is all there. He would know, for example, that he would not have to spend a great deal of effort and money on controlling the absolute level of the signal, but he had better be pretty careful about the signal-to-noise ratio. The point is that yesterday's research findings in a new field of human factors become tomorrow's standard laboratory practice. And everyday laboratory practice is the real rock on which a technology is built.

Let me now try to pick up Mr. Rathert's challenge about a taxonomy of decisions. The classification I will attempt to set up corresponds, in a rough way, to the time schedule he was talking about, or to how far ahead the decision is made. It is not a terribly good classification, because it depends on internal states of the subject that are not readily observable; but perhaps someone will think of a way around that.

One problem that psychologists have in thinking about classifying decisions—that is, psychologists of a certain stripe, who are heavily represented here—is that they have trouble thinking about any behavior that does not involve a decision. We are by now so thoroughly contaminated by computers that even a tracking task seems to involve multitudes of very rapid decisions—"The stimulus is to the left; now it is to the right." If a computer were doing the task, the program would be full of tests: bigger than? smaller than? According to the way many of us now think, these tests would seem to be decisions, and so this almost reflexive be-

havior gets included in the world of decision-making.

Our first category of decisions, therefore, includes the decisions that are almost like conditioned reflexes. These are the little adjustments made by a pilot hovering a helicopter on a moderately bumpy day, or by an operator performing some other kind of steering task. It is almost as if such decisions were made in a satellite computer. The reason I say "satellite" is that these decisions are not available to consciousness once the behavior becomes very skilled. (If any purists in the room wince at the word "consciousness," they may translate it. They can say these decisions do not leave traces that are recoverable in short-term memory.) It is as if consciousness were located in a central machine and these decisions were made in some sort of a satellite machine. The central machine can load a steering program into the satellite, but once the program is loaded and running, its internal workings do not leave any records that are available to the main program in the central machine.

One observes this sort of thing happening in flying very tight parade formation. It can be amusing to look out of the corner of your eye and see your hand moving, making small adjustments, and still be perfectly aware that the decisions on which those small adjustments are based are in no sense available to consciousness. You can see a correlation between your hand movement and the motion of the airplane you are tracking, but you cannot find the decisions by introspecting.[2] Here, then, is one category of decisions, an almost reflexive sort of behavior in which the decisionmaking is not available to consciousness.

The second large category I would propose is the category of pattern recognition or, to use the term more traditional in psychology, "concept identification." These are the cases in which some pattern of stimuli appears to a person and he identifies it as a "whosis" rather than a "zoosis." For example, at

breakout on a low approach, a pilot may be confronted with an almost infinite variety of visual patterns, but he has to make what is essentially a two-category decision. He classifies the given pattern into a category such as "continue to attempt visual approach," or "get the heck out of here."

This sort of decision covers a very wide range on Mr. Rathert's time scale (which is, perhaps, another difficulty in my proposed taxonomy). Down at the very short end, this kind of decision arises in steering a vehicle; that is, in the pattern identification of a pitchup or some of the other unstable conditions that he was talking about. It is the kind of thing that occurs when you are steering a vehicle and suddenly realize, "This thing is oscillating and my present steering program is only making matters worse." You switch quite consciously to introducing another steering program into the satellite computer, and leave that other steering program in effect until the oscillations have been resolved. In sum, on the short end of the time scale, this pattern-recognition type of decision results in a switch of program in the satellite.

On the long end, there is the sort of thing we are more accustomed to calling pattern recognition. One example I think of concerns a fellow who once came upon someone lying unconscious and said he could almost hear the voice of his old Red Cross first-aid instructor reciting, "When you see an unconscious person, look for bleeding, breathing, poison, shock, in that order." Here was pattern recognition: a pattern of stimuli that instead of triggering an immediate switch to another mode of motor behavior, triggered the execution of another program in the main computer. In this case it was a verbal program of the type that we would usually call a standard operating procedure (SOP). As I say, this second major category, conscious pattern recognition, spreads very, very wide on Mr. Rathert's time scale.

Finally there is a third category, which I would call explicit weighing and balancing. These are what the man in the street usually thinks of as decisions. They are the decisions

---

[2] Mr. Jeffress suggested that the original learning must have involved contemplation. Mr. Yntema agreed.

typified by, "Shall I hang around here over field X and assume that the thunderstorm is going to move off before my fuel gets critical, or shall I go to field Y, where the weather is quite good but my fuel will be very critical before I get down?" Introspectively, it seems that in making these decisions, a pilot switches his attention back and forth from one alternative to the other, as though he had in his head a program that was weighing one alternative against the other. This is very different from the second kind of decision. There, the process of making a decision consists, introspectively, of simply recognizing a pattern: once the recognition has been accomplished, the decision is as good as made. There is none of the conscious, explicit weighing and balancing that goes on in this third category.

To improve decisionmaking, we often try to move decisions from the third category into the second, and from the second into the first. I suspect that if you get to be very well trained on switch of control modes, the switch may become unconscious,[3] which means that training has moved the decision from the second category to the first. Similarly, operating personnel are often trained to have an SOP so that they will not sit there weighing and balancing. The high-school coach tells his quarterbacks: "On the fourth down, if you are inside such and such a yard line, do not think! Punt!" He is trying to replace weighing and balancing by pattern recognition that triggers an SOP.

WARD EDWARDS: Some educators, however, would argue that their main function in life is to move decisionmaking in the opposite direction; that is, to replace the automatic responses of childhood with more reasoned responses.

YNTEMA: I certainly am on their side. I guess the constraints of an operational situation often shove us in a direction that would not be desirable in education.

On another occasion I have called this weighing and balancing kind of decision ponderation. This antique but respectable word is related to the old meaning of the

word "ponder," which literally means to weigh in the mind.

Inasmuch as our chairman has said it is all right to go off and talk about one's own work (and who can resist that opportunity), I would like to talk about some old work I did on this kind of decision, in particular, on the question of using a computer to make these weighing and balancing decisions.

As missions speed up and become more complex, there is obviously going to be more and more pressure to have computers make decisions that we would really rather not see computers make. It is almost inevitable. You do not have to get terribly dramatic and think of a computer making big, crucial decisions that we want to reserve for the human. In most systems, there are a lot of little decisions that do not have immediate effects on the success of the mission or the safety of the system, but decisions that you would nonetheless like to have made with a certain leaven of human judgment thrown in. One way to make these decisions with some better approximation to correctness is to find a way to read out of the human expert the rules by which he would make decisions, and read the rules out in such a form that they could be read into a computer. Then the computer would make the decision pretty much as the human would.

The easiest way to explain all this is to take a practical example. A few years ago I helped some people at the Mitre Corp. in their test of a computerized air-traffic control system. A more complete description of all of the air-traffic control work discussed here is available (ref. 1). They were testing two versions of the computerized system, and were concerned that the tests were so complicated that the results were going to have to be scored by a computer. But they did not know how to set down logical rules by which a computer could take snapshots of the state of the system and arrive at a score of how well the system was functioning. They felt that the only way to decide whether an air-traffic control system was functioning well was to ask experts on air-traffic control. The experts they chose to ask were ATC con-

---

[3] Mr. Elkind indicated that he agreed.

trollers, all 16 of whom had at least 5 years' field experience and plenty of experience running this experimental system—presumably the best human judgment around. How could you read out from these people a judgment about what was a well-functioning system, getting the judgment in a form that a computer could use to score the results of tests?

These controllers agreed on 12 kinds of troublesome conditions that they felt had to be considered (such as: track misidentified, two aircraft in violation of radar standards, two aircraft will be in violation of manual control standards in 5 minutes, and so on). They felt that if they knew how many aircraft were in each of these 12 kinds of trouble, they could give a fairly good judgment of how well the system was functioning.

The controllers were given 26 pasteboard chits, on each of which a state of the system was summarized; that is, the numbers of aircraft in various kinds of trouble were specified in a notation similar to what the controllers were used to seeing on the computer-generated displays. Several of the chits are shown in figure 1. The top one on the right says there are two aircraft now violating radar separation standards, and two aircraft that will begin violating manual separation standards in 1 minute, unless something is done in the meantime. The bottom chit on

the left shows that four aircraft will begin violating radar standards within 4 minutes unless something is done about them, that two aircraft are now violating manual separation standards, and that one aircraft is misidentified (in other words, the system is tracking the wrong aircraft, a rather serious business).

The controllers were asked to express their evaluations of these 26 situations by laying the 26 chits out along a line. Figure 2 shows how the typical controller laid them out. ("Typical" is used here to mean the layout that correlated best with the layouts of the individual controllers.) At the top is the situation for which he would penalize the system most heavily in scoring the tests; next comes the situation to which he would assign the next largest penalty; and so on.

Furthermore, the way the controller spaced the chits was to indicate the relative sizes of the penalties he wanted to assign. For example, look at the three chits I have marked with white circles: they are about equally spaced. We told the controller that when he put three chits at equal spacing, like those three, we would understand that the penalty he wanted to assign to the middle situation was the average of the penalties he wanted to assign to the other two situations. In other words, a system that allowed the middle situation to occur twice would be penalized as much as a system that allowed the top situation to occur once and the bottom situation to occur three times. In psychologists' terms, we were trying to get the judges to lay the chits out on an "interval scale" of penalty.

To get these judgments into a form that can be used by a machine, the simple and obvious thing to do is to assume that the penalties for the 12 trouble-conditions are additive. In other words, if we have a situation in which the numbers of aircraft in the various trouble conditions are n_1, n_2, n_3, . . . , n_{12}, the penalty to be assessed for that situation is simply

$$\sum_j n_j p_j$$
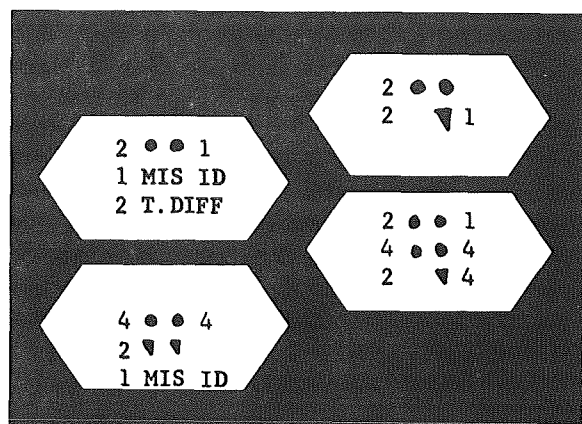
FIGURE 1.—Sample cardboard chits used in control simulation exercises.
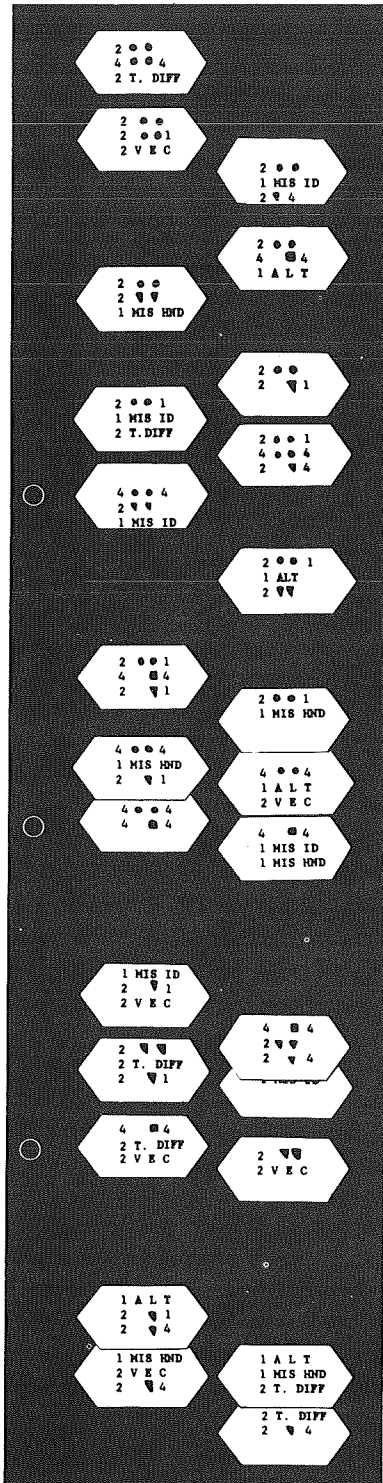
FIGURE 2.—Typical controller's evaluation layout of the simulation system in various states.

where p_j is the penalty to be imposed for each aircraft in the jth trouble condition.

Given the positions at which the typical controller put the 26 chits on the penalty scale, it is a straightforward problem in multiple linear regression to solve for the p_j. (And at the same time, to solve for the zero point on his penalty scale, the point that represents "no penalty." One has to solve for the zero point because an interval scale has no natural zero point.) Figure 3 shows the result. Each of the 26 points represents one of the chits. On the horizontal axis is plotted

$$S = \sum_j n_j p_j$$

where n_j is the number of aircraft that the chit says are in the jth trouble condition, and p_j is the value that came out of the regression analysis. The vertical axis shows the position at which the typical controller put the chit (the position being measured in arbitrary units from the zero point deduced from the regression analysis). If we had managed to describe the typical con-
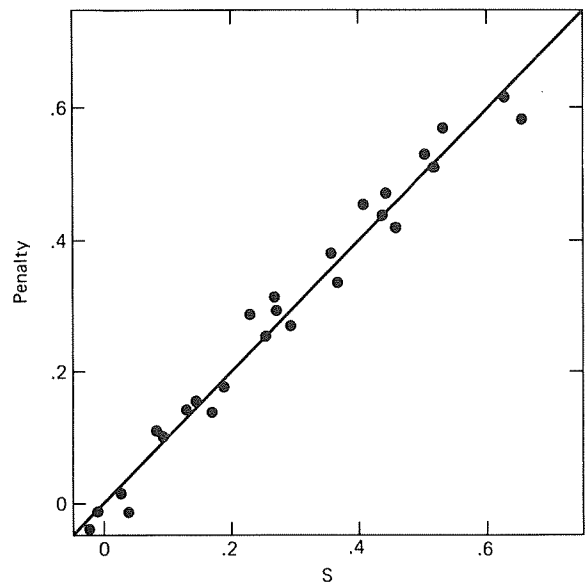
FIGURE 3.—Plot of where controller placed the chit in the layout in figure 2. (Each dot represents the height from the bottom of the board; line shows the computed penalty S.)

troller's judgments perfectly, all the points would fall on the straight line. The fit is not bad, which shows that the typical controller was pretty consistent in his judgments.[4]

The consistency reflects, perhaps, the fact that these controllers were professionals making a professional judgment about something that matters to them. Some of them took half a day at their desks to lay the chits out. They were not playing parlor games.

So now we have a mechanical way of scoring the system tests. Every minute or so the machine will take a snapshot of the air-traffic situation and will record the n_j, the numbers of aircraft in the various trouble conditions. The penalty assessed for that situation will be $\sum_j n_j p_j$, where the p_j are the 12 values that came out of the regression analysis.

A difficulty that arises in some judgmental applications of this kind is that judgments sometimes are not additive. Sometimes you cannot merely assign a penalty to each possible component of a situation and then compute a penalty for a whole situation by just adding up the penalties for components that appear in that situation. In other words, the total penalty may not be a linear function of the independent variables n_j. In that case you may have to add some higher order terms to the expression. Instead of approximating the overall penalty by $\sum_j n_j p_j$, you approximate it by

$$\sum_j n_j p_j + C_2(\sum_j n_j p_j)^2 + \cdots + C_\nu(\sum_j n_j p_j)^\nu$$

where the C's and p's are all parameters that must be adjusted to fit the data.

In the present case it turns out that although the fit in figure 3 was pretty good, it can be made a great deal better if we do use the squared term—that is, represent the judged penalty by an expression of the form

$$\sum_j n_j p_j + C_2(\sum_j n_j p_j)^2$$

Figure 4 shows the result. As in figure 3, the typical expert's judgment is on the vertical axis and S stands for $\sum_j n_j p_j$. The values of p_j used here are not, however, the same in fig. 3. The fit is noticeably better than in figure 3.
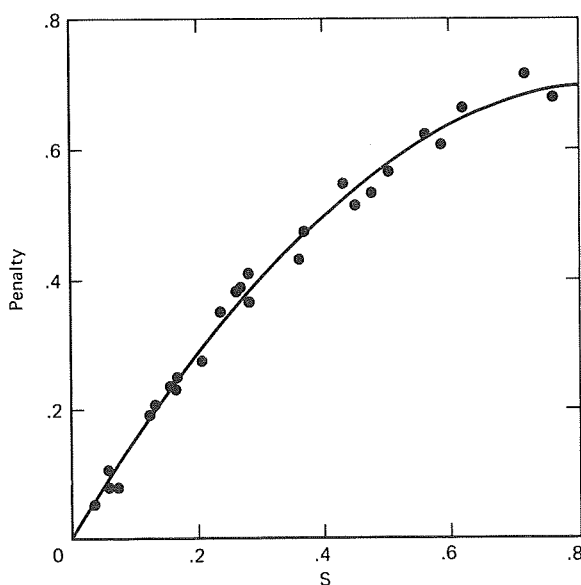
FIGURE 4.—Change of notation, using same values as in figure 3.

The difference between figures 3 and 4 is not, however, the main point. The main point is that in this application it proved possible to read out of human experts a rule that you could give to a machine to evaluate complex situations much as the experts would evaluate them. This, of course, is not yet a decision, but if you took two of these evaluations and had the machine choose the alternative that had the higher value, then that would be a decision.

Consider another example. Twenty Air Force pilots with a good deal of instrument

---

[4] It is mildly annoying that two of the situations turn out to receive negative penalties, but that is no cause for great alarm. That sort of thing can happen when a judge is not perfectly consistent.

time in the T–33 were asked to judge in what situations the average Air Force pilot with a standard instrument rating would be safer in landing a T–33. A more complete description of this experiment and of the technique for instructing the computer to refuse to make some of the decisions is available (ref. 2). The situations were defined in terms of ceiling, visibility, and fuel that would remain at touchdown if the penetration and approach were flown according to the book.

TABLE I.—*Sample from Book of Conditions Used for Judgment Exercises*

Condition	Situation A	Situation B
Ceiling, ft.	5000	1000
Visibility, mile	2	1
Fuel, gal.	15	200

Table I shows a page of a booklet containing 40 pairs of these situations, one pair to a page. The man was to look through the pages and put a checkmark under the situation that he thought would be safer for the average Air Force pilot. These 20 pilots were then asked to follow a procedure similar to the one used by the controllers. The procedure used with these pilots was a little different and, I now think, not as good; so I will not describe it. Nevertheless, each pilot was asked to lay out some chits in a way that would give the machine a rule for assigning a safety value to any given combination of ceiling, visibility, and fuel. Following this rule, the machine went through the same 40 choices that the man had made in the first part of the experiment, and we compared its decisions with his. This was done for each pilot, comparing the decisions he had made with the decisions made when the machine followed the evaluation rule that we obtained from his chits. In only 11 percent of the decisions did the machine fail to make the choice that the man had made.

This result looks pretty good if we compare it with the way the pilots disagreed with each other. The 20 pilots actually went through the experiment in pairs, so that the 2 men got exactly the same decisions to make before they were asked to express their rules for making the judgments. On the average, the two men disagreed with each other in 14 percent of the decisions. The difference between 11 and 14 percent is not statistically significant, but it does tend to suggest that the agreement between the machine and the man whose rule it was attempting to mimic was, if anything, better than the agreement between two men who were both competent judges of instrument flying conditions.

The decisions on which the machine disagreed with the man tended, as one would suspect, to be decisions that were not terribly disastrous; that is, cases in which the safety values that the man would assign to the two alternatives were evidently very close. If you crank that factor into the amount of disagreement, you can get a good measure of the seriousness of the machine's failures to mimic the man's choices.

That measure can be used to investigate the tactic of telling the machine to refuse to make certain decisions: it hands them back to the man and tells him to make them himself. There are, of course, certain classes of decisions that the machine makes well, and others that it makes poorly; and it turns out that the machine itself can be instructed to compute a rough index of the probability that it will make a serious error on any particular decision. If the index exceeds a certain threshold, the machine refuses to make the decision and tells the man to make it himself.

For example, in the present experiment you can set the threshold in such a way that the measure of the seriousness of the machine's failures would be improved by a factor of 3, at the expense of burdening the man with 19 percent of the decisions. In a practical application, you would, of course, pick the threshold according to how critical the decisions were going to be or, conversely, how much time the man could spare from his other duties.

DISCUSSION

EDWARDS: In all of these instances that you have given us, there is a well-defined dimensional analysis. Furthermore, the dimensions are common to all stimuli of the class you expect to be dealing with. I find many cases coming my way these days in which that just is not so. There may be obvious dimensional analysis of one stimulus, but other stimuli that are going to have to be included in the same utility space have a different set of dimensions. I do not see any way of doing anything more sophisticated than just plain asking for an evaluation of each stimulus that comes along, do you?

YNTEMA: No; I do not. Unless you can describe the situation in terms of some fairly clean, quantifiable dimensions, it is going to be hard to tell the machine what the situation is. To tell the machine, "Here is the situation," you pretty well have to have the situation broken down into dimensions that are coded into so many bits—unless you get into pattern recognition of an advanced sort, which we would all like to see machines get to some day.

EDWARDS: I would add, as I am sure you would agree, that the notion of getting human beings to judge utilities is in no sense inconsistent with the notion of having the machine make the decisions.

REFERENCES

1. MOREY, J. L.; AND YNTEMA, D. B.: Experiments on Systems. *In:* Second Congress on the Information System Sciences (J. Spiegel and D. E. Walker, eds.), Spartan Press (Washington), 1965.

2. YNTEMA, D. B.; AND KLEM, L.: Telling a Computer How To Evaluate Multidimensional Situations. IEEE Trans. on Human Factors in Electronics, HFE–6, 1965, pp. 3–13.

Signal Recognition as an Analog to Decisionmaking in Limited Visibility Landing

TRIEVE A. TANNER, JR.

NASA Ames Research Center

The purpose of this paper is to suggest how some decisionmaking situations in which pilots are involved may be viewed as being analogous to laboratory tasks in the area of signal detection and recognition. The research that has been conducted in this area has produced a considerable amount of information about the performance of human observers in the laboratory. Accompanying this research has been the development of mathematical models of detection, recognition, and related behavior. These models have been successful in accounting for performance under a variety of laboratory conditions. To the extent that these laboratory tasks are analogous to decisionmaking situations of a pilot, the results of the research and the related theoretical developments may be valuable for the description, prediction, and, through training, the improvement of a pilot's performance.

Discussions of the theoretical developments in signal detection and recognition are readily available elsewhere (refs. 1 and 2). Rather than discussing theory, it is my intent to indicate how some of the findings from recent signal-recognition studies may have implications for decisionmaking by a pilot.

In one of the simpler forms of a detection or a recognition task, one of two stimulus events is presented to an observer and he is asked to judge which event has occurred. The two events are selected so as to be very difficult for the observer to discriminate.

Any sense modality may be involved, although most of the research has been done with either visual or auditory stimuli. In signal detection, one of the events consists of a faint signal that is embedded in background noise, and the other event consists of noise alone. In the simplest form of the recognition task, the stimulus events are two signals that are easily detectable, but are adjusted so as to be difficult to discriminate.

Several situations faced by a pilot may be considered as involving a choice between two alternative decisions in a way that is similar to these laboratory tasks. One example of such a situation is involved in the decision whether or not to eject from a fighter aircraft (ref. 3). My remarks will be directed to another example: the problem of landing under conditions of limited visibility; that is, landing through low overcast weather conditions (ref. 4). My description of the landing situation is obviously an oversimplification of the problems that are involved. However, the purpose here is not to provide a complete description of the performance skills involved in landing an aircraft. I am concerned merely with one act of the pilot: his decision whether or not to land.

In the situation of interest, the pilot directs the descent of the aircraft through the clouds by reading guidance instrumentation. At a critical altitude and distance from the

end of the runway, he must decide not to land but to go around and, perhaps, make another approach, if he has not broken out of the clouds. The actual values of these critical distances vary with the handling qualities of the aircraft and the guidelines issued by governing bodies like the FAA and the U.S. Air Force, and are not important to this discussion. If breakout from the clouds occurs substantially farther from the runway and higher than what is considered to be critical, the pilot has ample time to adjust to any inaccurate information that might have been received from his instruments and to land using normal visual cues. If breakout occurs somewhere between these two points (i.e., between a mandatory missed approach and what is essentially a normal visual landing), the pilot must quickly decide whether or not he is "in the slot," that is, in proper alinement with the runway, on the proper glide slope, and so forth. If he is not in the slot, he must decide whether or not his altitude and distance from the runway will allow sufficient time for him to make the proper adjustments for a visual landing. If he decides that he does not have time to make such adjustments, he must go around.

There are a total of four possible outcomes from the pilot's two decisions. In figure 1, these outcomes are arranged in a 2 by 2 matrix, which represents the intersections

of the two-decision alternatives (to land or to go around) and the actual conditions, dichotomized into those in which a successful landing is possible and those in which it is not possible. If the pilot decides to land, he may land successfully, as represented by cell A, or unsuccessfully, cell C. If he decides to go around, he may be correct, cell D (i.e., he really could not have landed successfully), or he may have executed an unnecessary missed approach and really could have landed successfully, cell B.

The distinction between a successful and an unsuccessful landing, cells A and C, can be verified objectively, at least in principle, depending only on the definition of "successful." However, the distinction between a correct and an incorrect go-around is impossible to measure objectively. This distinction depends on a judgment (made either by the pilot or some other observer of the landing conditions) as to whether or not a landing was possible. The pilot himself may believe that he is seldom, if ever, wrong when he decides to go around, because, at the moment he makes his decision, he is the best judge of the situation and of his own capabilities to meet it. His superiors, however, may at times be critical and suggest that he was overly cautious. I assume that pilots as well as executives would agree that sometimes a pilot may be overly cautious in executing a missed approach, just as he may sometimes be overly adventurous in deciding to land.

The pilot's decision to land or to go around obviously is influenced by his sensory capacity to discriminate how far he is from being in the slot. A pilot who is highly accurate in evaluating his position at breakout and his ability to correct it if necessary will, over a long series of landings, have a high number of outcomes in cells A and D (fig. 1) relative to the number in cells B and C. This situation is shown in the upper matrix of figure 2 (labeled "Higher Sensitivity"). The lower matrix shows the situation for a less-accurate or less-sensitive pilot. However, the pilot's decision is influenced also by whatever factors would

Decision Alternatives

	Land	Go Around
Successful Landing Possible	Successful Landing A	Unnecessary Missed Approach B
Not Possible	Unsuccessful Landing C	Correct Missed Approach D

Actual Conditions For Successful Landing

FIGURE 1.—Decision matrix for low-visibility landings.

FIGURE 2 — Higher Sensitivity

Decision Alternatives

Actual Conditions For Successful Landing	Land	Go Around
Possible	Many — A	Few — B
Not Possible	Few — C	Many — D

FIGURE 2 — Lower Sensitivity

Actual Conditions For Successful Landing	Land	Go Around
Possible	Few — A	Many — B
Not Possible	Many — C	Few — D

FIGURE 2.—Decision matrices for pilots with relatively high and low sensitivity.

FIGURE 3 — Greater Caution

Decision Alternatives

Actual Conditions For Successful Landing	Land	Go Around
Possible	Few — A	Many — B
Not Possible	Few — C	Many — D

FIGURE 3 — Less Caution

Actual Conditions For Successful Landing	Land	Go Around
Possible	Many — A	Few — B
Not Possible	Many — C	Few — D

FIGURE 3.—Decision matrices for pilots with relatively high and low caution.

cause him to be more or less cautious. Motivation is perhaps the most obvious factor that will influence the degree of his caution. Motivation to maintain his own safety and that of the passengers will tend to increase caution, whereas an attempt to maintain prestige or to save the time and cost involved in a missed approach will tend to reduce caution. These situations are shown in figure 3. A greater degree of caution (shown in the upper matrix) is represented by higher frequencies of occurrence in cells B and D than in cells A and C. A lower degree of caution (shown in the lower matrix) is represented by relatively higher frequencies in cells A and C. The results of laboratory research on signal detection and recognition suggest that any factor that might influence a pilot's expectation of experiencing successful landing conditions at breakout will, in turn, influence his decision toward greater or less caution. The relative frequency with which the pilot experiences successful and unsuccessful landing conditions and the amount of information feedback that he receives from his decisions may be such factors. These will be discussed in more detail in relation to the laboratory research.

Data to fill the decision matrix in figure 1 are not available from actual landing approaches. It is clear that an abundance of entries may be obtained for cell A (successful landings). However, because the occurrence of an unsuccessful landing is likely to result in fatality, only one entry in cell C might be obtained for a given pilot. Even relatively unsuccessful landings that are short of being fatal to the pilot are, fortunately, infrequent. As noted previously, outcomes that would fall into cells B and D (correct and incorrect missed approaches)

cannot be distinguished in the actual flight situation. If we could obtain, from a particular pilot, a substantial amount of data in each of the four cells of the performance matrix, then we could compare his performance with that of observers in the laboratory situation, and we could determine the extent to which the two are comparable. The use of an aircraft landing simulator would allow the collection of the desired data. By defining objective criteria for acceptable and nonacceptable landing conditions, data in all four cells of the matrix could be obtained. I shall describe a plan to initiate such a simulation program, after discussing some of the laboratory research to which I have referred.

In the type of signal-recognition study that is of interest here (refs. 5 and 6), the two signals that an observer is asked to discriminate are two 1000-Hz tones that differ from each other only in amplitude. These will be referred to as the loud and the soft tone. One of the two tones is presented to the observer on each trial of a randomized sequence of trials. The number of trials in a session and the number of sessions vary according to the number of parameters and the number of values of the parameters that are investigated. The observer's task is to judge which tone is presented on each trial. His response is a simple button press. Both tonal amplitudes are clearly audible, but they are adjusted over a series of preliminary trials so that an observer performs at a level of about 75 percent correct responses; that is, about halfway between chance and perfect accuracy. As with the pilot's decision whether or not to land, there are four possible outcomes for a trial. These outcomes are shown in figure 4. After a loud signal occurs, the observer can say that the loud tone has occurred, cell A, or that the soft tone has occurred, cell B. Similarly, after a soft signal occurs he can judge that the loud tone, cell C, or the soft tone, cell D, has occurred.

Two variables that have been investigated in these studies are the proportion of trials on which each of the signals is presented

FIGURE 4.—Decision matrix for two-tone recognition task.

to the observer, and whether or not the observer receives information feedback after his decision is made on each trial. In the two experiments cited below, both conditions—that is, the proportions of loud- and soft-tone trials and the presence or absence of information feedback—were held constant over each session.

In a study conducted by Kinchla (ref. 5), the loud tone was presented on either 0.25, 0.50, or 0.75 of the trials of a given session. In each case, the soft tone occurred on the remainder of the trials. Following the observer's decision of whether the loud or the soft tone had occurred, he was told, with an information light, which tone actually had occurred on that trial. In addition to this information feedback, the observer was told at the start of a session whether to expect 0.25, 0.50, or 0.75 of the trials to contain loud (or soft) signals.

Data collected in this study are presented in figure 5. The ordinate in this figure represents proportions of responses. The abscissa represents the proportion of trials on which a loud signal occurred. The upper curve represents the proportion of loud signals that observers correctly judged as loud. The lower curve represents the proportion of soft signals that were incorrectly judged as loud. Therefore, these curves plot the proportions of outcomes in cells A and C of figure 4, contingent on the signal that occurred. These proportions usually are referred to as con-
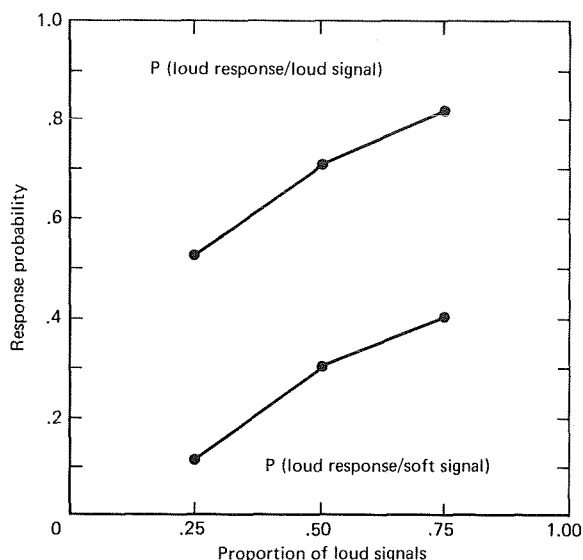
FIGURE 5.—Contingent probabilities for a loud response with information feedback.

tingent probabilities. Thus the upper curve may be called the probability of a loud response given a loud signal, $Pr(LR|LS)$, and the lower curve may be called the probability of a loud response given a soft signal, $Pr(LR|SS)$. It is unnecessary to plot the contingent probabilities for cells B and D because they are complementary to the contingent probabilities of cells A and C, respectively.

Figure 5 illustrates that as the probability of a loud signal increased, both the probability of a loud response given a loud signal and the probability of a loud response given a soft signal increased. It should be noted that the changes in these contingent probabilities were independent of the accuracy of the observer's decisions; the total proportion of correct responses did not change appreciably as the signal probability increased. Because a simultaneous increase in both contingent probabilities for a loud response is independent of accuracy, it is often referred to as a response bias. Hence, in Kinchla's study (ref. 5), there was an increasing bias for the loud response, as the proportion of loud signals increased. These results are consistent with the results of signal-detection studies (ref. 2).

A study by Tanner et. al. (ref. 6) was similar to Kinchla's in that no information feedback was given to the observers and at no time were the observers told what signal proportions to expect. In this experiment, five signal proportions were investigated. For any given session, the observer was presented the loud signal on either 0.1, 0.3, 0.5, 0.7, or 0.9 of the trials. Again the signal proportions are complementary, so that the soft signal occurred, respectively, on 0.9, 0.7, 0.5, 0.3, or 0.1 of the trials.
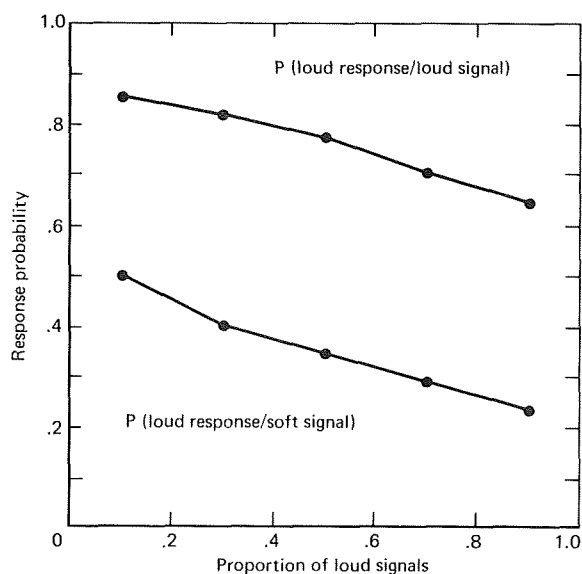
FIGURE 6.—Contingent probabilities for a loud response with no information feedback.

Data from this study are presented in figure 6, where the decision probabilities are plotted against the five possible proportions of loud signals. As in figure 5, the upper curve represents the probability of a loud response given a loud signal and the lower curve represents the probability of a loud response given a soft signal. Under this condition of no information feedback, the trends in these curves are the inverse of those shown in figure 5. Rather than increasing, as in Kinchla's study (ref. 5), the bias for the loud response decreased, as the proportion of loud signals increased.

Subsequent   experiments   (unpublished)

have confirmed the results of these two studies and, in addition, have given evidence that the biases reported by Tanner (ref. 6) are not dependent on the observers being completely ignorant of the fact that the signal proportions were being varied, as were the observers of Tanner (ref. 6).

Thus, when observers in these laboratory recognition studies had information regarding the actual signal proportions, an increase in the proportion of the loud signals resulted in an increased bias toward judgment that the loud signal had occurred. However, whenever information about the signal proportions was eliminated, an increase in the proportion of loud signals resulted in a decreased bias toward the decision that the loud signal had occurred.

Another finding of these experiments concerns the way in which a decision is influenced by the observer's responses on previous trials. It has been found that when the observer receives no feedback, his bias toward reporting either the loud or the soft signal is high (when he reported that the same signal occurred on the immediately preceding trial). In other words, responses tended to occur in runs. When feedback was given, previous responses had much less influence on current responses.

Assuming a pilot in a limited-visibility landing approach to be similarly influenced by the interaction of these two variables (feedback and the proportions of stimulus events), then he would be biased toward landing if he were accustomed to experiencing successful landing conditions at breakout and if he received information feedback after every landing approach. Also, if he were accustomed to unsuccessful conditions with no feedback, we would expect him to be biased toward landing. If he were accustomed to acceptable conditions with no feedback or to unacceptable conditions with feedback, then he would be biased toward executing a missed approach.

Obviously, these proposed relations are hypothetical. Because a pilot receives feedback only when he lands, predictions of his decision are meaningful only when they are based on previous landing with feedback or previous missed approaches with no feedback. The above predictions take no account of the decision that was actually made by the pilot on previous approaches, but are based only on the stimulus conditions. If predictions are based on the effects of previous performance found in the recognition studies that have been cited, then, independent of the stimulus conditions, the pilot might be expected to repeat the decision that he made on his previous approach. However, because this prediction is based on no feedback being received, it is relevant only to missed approaches.

Therefore, extrapolating from the results of previous laboratory recognition research to predict a pilot's bias toward more or less caution in landing must be considered highly speculative. The situation in the laboratory could be made more similar to the conditions of landing if the presentation of information feedback were made contingent on the observer's decision. Such an experiment is currently being conducted in the Human Performance Laboratory of the Ames Research Center. The task—recognition of two auditory amplitudes—is essentially the same as that investigated by Kinchla (ref. 5) and by Tanner (ref. 6), except that feedback is given following only one of the observer's two possible decisions. That is, half of the observers are given feedback whenever they judge that the soft tone has occurred; the other half are given feedback only when they judge that the loud tone has occurred. The purpose of the experiment is to determine whether or not response-contingent feedback with variations in the signal proportions produces decision biases that are similar to those found in previous investigations. Whether the biases are like those found with 100 percent feedback (ref. 5), with no feedback (ref. 6), or if the effects are different from either of these should have implications for decisionmaking in the limited-visibility landing situation. Therefore, the results of this study will be used in assessing the feasibility of a more direct investigation of the landing problem. If the

contingent-feedback situation does produce some type of decision bias, as expected, the results will be used in determining a strategy for applying the recognition paradigm to research on an aircraft landing simulator.

In an actual landing situation or the simulation thereof, the views of a runway that a pilot experiences at breakout are considerably more complex than the two tones that have been used in the recognition studies. These views may vary along several dimensions, depending on such factors as the altitude, distance from the runway, and attitude of the aircraft. In addition, the values along each of these dimensions are continuous rather than discrete, as are the two tones in the auditory recognition task. Thus, on a given approach, a pilot may experience any one of an infinite set of possible views of the runway. Whether or not a dichotomization of these views, into acceptable and unacceptable for successful landing, would yield data similar to that obtained in two-tone auditory recognition studies could be determined using a landing simulator. As suggested previously, recognition of acceptable and unacceptable landing views analyzed similarly to the simpler two-signal recognition task may result in a better understanding of the pilot's decision whether or not to land after breakout.

DISCUSSION

WARD EDWARDS: I am puzzled by the apparent inconsistency between the data reported in figure 6 and the typical gambler's-fallacy data that you get in either binary prediction or binary production. Usually you are likely to get too many alternations rather than too many repetitions.

TRIEVE A. TANNER: That is not what we found.

EDWARDS: I know. I am wondering if you have any idea why your data do not come out in that grand and ugly tradition.

TANNER: I can refer you to a model developed by Haller and Atkinson, which attempts to explain the various results of this study (ref. 6).

RICHARD C. ATKINSON: A comment here, Ward. In that experiment (fig. 6), the subject never realizes from day to day that the proportion of loud and soft signals has been changed. His feeling about the task is that it is somewhat more difficult on some days than on other days, but the subjects did not report realizing that the signal proportions change from day to day.

TANNER: That is true. I failed to mention that. Subjects actually have reported that they thought we made the task more difficult from day to day.

ATKINSON: What it amounts to is that the task ceases to be one of signal detection and becomes one of memory. You make a judgment of a given input not on the basis of comparison with background noise, but by comparing it with the last input and recalling how you judged that last input. When you have feedback on the situation so that the subject knows precisely how that last judgment should have been made, quite a different performance record is created as opposed to when the subject has only his estimate of how the judgment should have been made.

EDWARDS: Feedback also sets up a well-defined prior probability.

ATKINSON: If in the first task you plotted just the probability of a loud tone, it would have virtually matched; but in the second task, if you had plotted the probability of saying loud, it would have been about 50 percent, that is to say, independent.

TANNER: In fact, it actually increased slightly.

EDWARDS: Does that indicate that subjects would find the tasks easiest when loud and soft stimuli were each occurring about one-half of the time?

TANNER: That seemed to be the case.

EDWARDS: So the nature of the difficulty in judgment is simply that subjects are committed to calling some loud stimuli soft if there are too many loud and some soft stimuli loud if there are too many soft.

LLOYD A. JEFFRESS: Are the data in figure 6 taken with the subjects explicitly or implicitly believing that the a priori probability is 0.5?

TANNER: They are simply not told.

ATKINSON: In a more recent study, Haller (ref. 7) actually took a careful protocol of subjective experience at the end of the experiment; to my knowledge, no one ever reported having realized that the signal probabilities were changing in proportion.

JEFFRESS: Another possible approach here would be to tell the subject that the a priori probability may be anywhere from 0.1 to 0.9, that it will remain the same during one sitting, but is likely to change without notice.

TANNER: In fact, the third study, which I briefly mentioned, used subjects who were not naive with respect to the a priori signal probabilities. We made them not naive by running them in a feedback condition, but it took them some time. I do not have the final results yet, but it looks as though the ROC curves flipped upside down again. Subjects were put back in the no-feedback condition after a long series of sessions.

ATKINSON: For people who are interested in signal detectability, let me comment on the model. It is

very simple. It assumes that you have some memory of the last signal that was presented and that you have your current input. You just compute a difference on the two. You look at that difference. If the difference is highly positive, then what you have now is far above what you had last time, so you tend to call it loud. If the difference is highly negative, you tend to call it soft. If it is in the midrange, you have a problem of what to do. When there is no information feedback, the assumption is that you call the signal the same thing you called it last time. When there is information feedback, you call it what the experimenter called it the last time. That sort of model will predict the change in performance from the feedback to the no-feedback condition. The model also predicts that sequential effects are mammoth in the no-feedback case, and more nearly minimal in the feedback case. So, it is a signal detectability model, but we have two criteria mapped onto the difference scores—the difference scores being defined in terms of what you have now and what you remember from the last run. It fits very well with the subjects' subjective experience. One way of viewing this is that over a whole host of trials, the subject builds up some memory that he is always comparing against. Another way of viewing it is that the subject is really just keeping track of the last signal and comparing with that. That is what subjects tell you they tend to do. They do not have some long average memory, but they tend to relate to the last signal.

JOSEPH MARKOWITZ: Did you give the subjects feedback at the end of the run or the end of the day?

TANNER: We have done both.

MARKOWITZ: The reason I ask is this: sometimes, when you go to the asymmetric probabilities, subjects tell you it is harder even in the no-feedback case. So they are recognizing that they have inappropriate criteria, but are not changing the criteria —which I am not sure I understand.

TANNER: Even when subjects are told how well they are doing, it is just that the proportion of correct responses does not change much.

ATKINSON: Is that true?

TANNER: Yes; not enough so that they notice much difference. I am sure it must be noise to them.

MARKOWITZ: Tell me why they say it is harder with the asymmetric probabilities.

ATKINSON: It is harder because frequently the same stimulus is being presented again.

MARKOWITZ: So they simply have many of these uncertainty decisions, is that the idea?

TANNER: Yes.

JEFFRESS: Carried to the extreme, if you had 100 percent loud signals, you would have a time trying to decide which ones were soft.

MARKOWITZ: Right. In the feedback case it is easy.

EDWARDS: Have you explored asking them afterwards in the no-feedback case to estimate the proportion of loud stimuli?

TANNER: No; we have not.

EDWARDS: Actually you could play that two ways. You could ask them to estimate the proportion of loud stimuli and the proportion of loud responses. Chances are they would do well in estimating their own behavior, but it would be interesting to see if they have such a strong bias in favor of a 50–50 a priori probability. Obviously, you should be able to influence this by influencing discriminability of the stimuli. It becomes essentially a random-sequence production task if the stimuli are completely indiscriminable. It becomes a relative frequency estimation task if the stimuli are perfectly discriminable. It moves gradually from the one kind of task to the other.

TANNER: I think it may be time to ask the subjects to estimate these proportions. We did not want to allude to signal probabilities in the first task at all because we did not want to bring these to their attention.

REFERENCES

1. SWETS, J. A.: Signal Detection and Recognition by Human Observers. John Wiley & Sons, Inc., 1964.
2. GREEN, D. M.; AND SWETS, J. A.: Signal Detection Theory and Psychophysics. John Wiley & Sons, Inc., 1966.
3. CHUBB, R. M.; BRAUE, G. C.; AND SHANNON, R. H.: Ejection Capability versus the Decision to Eject. Aerospace Med., vol. 38, 1967, pp. 900–904.
4. BECK, R. H.: The Revival of the One Man Band. The Airline Pilot, vol. 36, 1967, pp. 7–21.
5. KINCHLA, R. A.: A Comparison of Sequential Effects in Detection and Recognition. Tech. Rept. 1, Psychol. Dept. Exptl. Psychol. Ser., N.Y. Univ., 1966.
6. TANNER, T. A.; HALLER, R. W.; AND ATKINSON, R. C.: Signal Recognition as Influenced by Presentation Schedules. Perception & Psychophys., vol. 2, 1967, pp. 349–358.
7. HALLER, R. W.: The Role of Information Feedback in Signal Recognition. Doctoral dissertation, Stanford Univ. Ann Arbor, Mich.: University Microfilms, 1969. No. 69–13960.

Attention[1]

ALFRED B. KRISTOFFERSON

McMaster University

This is a review of the research on attention, time, and human information processing that has been carried out in my laboratory during the past 4 years. For the purposes of this conference, I will stress the concepts that are involved and their coordination to observables, and major conclusions from the experiments rather than detailed discussions of data. It would also seem to be appropriate for me to indicate generality when that is possible and to tell you what I know about the limits that must be imposed upon that generality.

The term "attention" has a number of very different meanings. Historically, there have been two major meanings; one implying selectivity in perception and the other emphasizing perceptual clarity. During the past 10 years or so, a number of new meanings have accrued to the term as a result of the recent surge of research in the area, and we seem to have reached a point where it is hardly useful to continue to use the single term. My work does not attempt to systematize this entire field. Instead, I am concerned with working out the details of one particular attention mechanism—one that may be described as a mechanism that controls the flow of sensory information by selecting among sensory inputs.

My approach is an indirect one in that I make inferences about this selective mechanism through measurements of certain of its temporal characteristics. A concept has evolved out of this, linking the present theory to a rather large body of theories that seem to have little or nothing to do with attention. I am referring to the concept of a time quantum or a unit of psychophysical time. At present, the time-quantum hypothesis is a central part of this theory and I will show how it is related to attention and to central information processing.

Two relevant reviews have appeared in recent literature. In 1963, White reviewed the many theoretical and experimental studies that bear upon the concept of a psychological unit of duration (ref. 1). More recently, a rather different review has been presented by Harter (ref. 2), a review of the various hypotheses of central intermittency in perception. Harter discussed seven different classes of experiments that have provided support for the idea of central intermittency and also the various theories of central intermittency that have been proposed. The present theory has some elements in common with most of the theories that Harter analyzes.

EMPIRICAL PARAMETERS AND SOME EXPERIMENTAL RESULTS

This project began with a specific statement of a very old, largely philosophical proposition concerning attention. It is the idea that attention is limited in that one can pay attention to only a small range of sen-

sory inputs at any one moment. The extreme form of this proposition is that one can attend to only a single input at any particular point in time and that attending to multiple inputs requires switching of attention from one to the other. Therefore, I thought that I could begin by setting up the working hypothesis that one can pay attention to only one thing at a time and that if one is paying attention to thing A at the time thing B occurs, it should require some length of time to switch attention from A to B. It seemed feasible to attempt to measure this switching time.

There are, however, immediate difficulties that arise when one sets out to do this, and very careful experimental procedures are required to surmount these difficulties. Some of these difficulties are inherent in the very simple notion that I have just described. For example, if one wishes to measure the switching time between A and B, it is necessary to be sure that a subject is, in fact, paying attention to A at the time B occurs. My experiments lead me to believe that one cannot safely assume that this is the case; under some conditions it may be the case, but usually a more detailed theoretical model is required to take into account the fact that ordinarily subjects cannot meet this criterion. Another requirement is that the switching itself must be accomplished reliably; that is, that attention go from A directly to B rather than via some intermediate channel. This calls for the use of experimental arrangements that are clearly defined for the subject and for adequate practice sessions.

Because the switching time is undoubtedly very short, a third requirement is set upon attempts to measure it. This is that the times of occurrence of the sensory inputs must be specifiable with a rather high degree of precision. These considerations and others led me to design experiments using very simple sensory signals. In all the experiments, the signals are spots of light and pure tones, both of which can be controlled precisely along the time dimension. I use lights and sounds to be as certain as is possible to be that the signals really are different things

that cannot be attended to simultaneously. And, finally, the critical signal events are the offsets of lights and sounds rather than their onsets, in order that the locations of A and B will be well defined for the experimental subject so that he will be able to attend reliably and switch his attention reliably from A to B.

These considerations are not very systematic and I have not presented them at all fully, but they do give some indication of the reasons why I have done experiments the way I have. Also, they introduce the general logic of the experiments. Two very different methods of measuring the switching time of attention have come out of this. The first one has to do with measuring the length of time required for a message to travel through the entire system, from signal to response. These are experiments on reaction time. The basic idea is that if one could measure the time required to respond to a signal B when the subject is in fact paying attention to B at the moment it occurs, and then measure the time required to respond to B when the subject is attending to A at the time B occurs, then by comparing these reaction times one should be able to infer the switching time between A and B. This is one approach, and one that can be applied only indirectly, as I will explain in a moment. The other approach requires the experimental subject to discriminate the relative time of occurrence of two different things. If he cannot pay attention to A and B simultaneously, than his ability to judge whether A and B occur simultaneously should be limited by the speed with which he can switch attention between them. This second approach has led us to do a large number of experiments on successiveness discrimination—the ability of human subjects to discriminate the successive occurrence of different sensory events from the simultaneous occurrence of such events.

I want to go on now and describe some specific experiments. I would like to avoid details as much as possible, but because many of the details of procedure are of critical importance, as I will discuss in a

later section, I hope I do not create the impression of greater generality than is warranted.

Successiveness Discrimination

One form of the successiveness discrimination function is shown in figure 1. It is a relationship between the probability of correctly discriminating the successiveness of a light and a sound and the time interval that elapses between the occurrence of the two signals. The time interval between signals is shown in figure 1 to extend in the positive direction from zero and it is so defined that these positive values mean that the visual signal occurs before the auditory signal. At zero, the objective time of occurrence of the two signals is simultaneous. And there is, of course, a negative side to this in which the auditory signal occurs first.
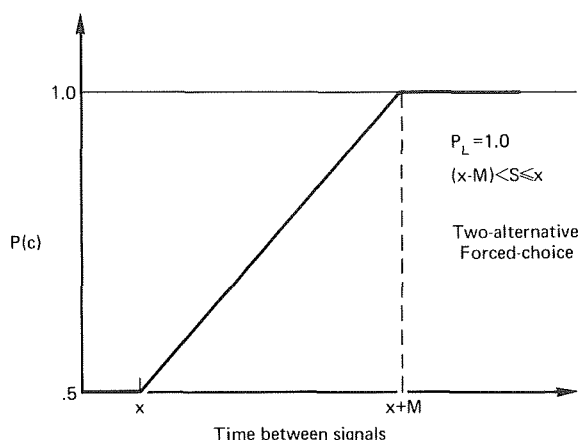
FIGURE 1.—Theoretical form of the successiveness-discrimination function.

The proportion of correct discriminations is plotted on the ordinate from 0.5 to 1.0 because a two-alternative, forced-choice method is used in which the subject is given two signal pairs on every trial and is required to decide which pair is the successive pair. This means that even when he is completely unable to discriminate between the pairs, the subject will still get one-half of the trials correct.

Figure 1 indicates that the theoretical

function that is shown is the one to be expected only when the time interval between the signals in the simultaneous pair on each trial is within a certain range. One cannot assume that a light and a sound that occur simultaneously will be simultaneous psychologically because the effects of the two signals are conducted at different rates over the two sensory systems. However, it follows theoretically, as I will explain later, that, as far as this discrimination is concerned, the two signals in the simultaneous pair will be exactly equivalent to a psychologically simultaneous pair, provided that they occur within a certain range of times of each other. This range, translated into real time, is indicated in figure 1.

The theoretical successiveness function that is drawn in figure 1 is a linear function that has three parameters. One of these parameters is P_L, the probability with which the subject is paying attention to the visual channel at the moment the visual signal occurs. Figure 1 indicates that a function consisting of a single linear segment is to be expected only when P_L is equal to 1.0. Furthermore, the other constraint—that it is sufficient for the interval of the simultaneous pair to fall within the range between x and $(x-M)$—is satisfied only when P_L is equal to unity. In our early experiments, we tried to arrange experimental conditions so that P_L would meet this theoretical specification, and so that we could analyze the data by using this single linear segment.

The other two parameters are x and M. The first of these is the time separation between signals at which the function just begins to exceed the chance level. In theory, x is the interval between the signals at which the relevant neural effects produced by the signals occur simultaneously. Because afferent conduction time is slower in the visual channel, we expect x to be positive—and we find that it is. The final quantity is the one of special interest. The function ascends from chance to 1.0, over the range between x and $(x+M)$ milliseconds. If our interpretation of x is correct, then the span of the function, the quantity M, is the mini-

mum time that must separate the two independent neural events for them to be discriminated as successive with complete certainty. Stated empirically, M is the minimum time that must be added to the interval between the two signals to bring the probability of a correct response from chance to 1.0.

We have obtained successiveness discrimination functions that are quite well described by figure 1. We now know that figure 1 describes a limiting case that can be closely approached but not completely attained. The first experiment that I want to discuss does assume that the case described by figure 1 is actually achieved, but I will indicate later the extent to which that is not quite correct.

Effect of Channel Uncertainty Upon Discrimination Reaction Time

The time required for a subject to make a response to a stimulus, when he is instructed to respond as rapidly as he can, is very different from one occasion to another. The amount of variance in such reaction times is especially great when the subject must discriminate between signals in order to determine whether or not to make the response on each trial. The reaction times to which I refer here are called discrimination reaction times. The subject is confronted with multiple signals, but he is required to make, or not to make, a single response. In a typical experiment, three signals are presented on each trial, two visual and one auditory. At the end of a certain fixed length of time, one of the three signals goes off. The subject has been instructed to release his hand from the key as rapidly as he can if either the tone or the right light goes off, but to withhold making the response if the left light goes off. At the beginning of each trial, a cuing signal is given. This signal tells the subject that if the stimulus is a positive one it will be in the visual channel, or that it will be in the auditory channel, or that it may be in either of the two channels. The subject knows that on every trial the negative signal, the

left light, may occur and it does occur on one-fourth of the trials.

Two distributions of reaction times are obtained for the visual signal: one for those trials on which the subject is "certain" and one for the trials for which he is "uncertain"; the same is true for the auditory signal. Thus, four distributions of reaction times are obtained in this experiment.

We are interested in the effect of uncertainty because when the subject is uncertain there should be a larger proportion of trials on which he is paying attention to the wrong channel at the moment the signal to respond occurs. If uncertainty has only the effect of requiring the subject to switch attention on some trials, then, by looking at the increments of time that are added to reaction time as a result of uncertainty, we should be able to infer something about the switching time of attention. One way of accomplishing this is outlined in figure 2.

For a particular signal, either the visual or the auditory, we have the measurements that are shown at the top of figure 2. There is a mean reaction time and a variance for the case when the subject is certain and

Measurements		
Condition	Mean	Variance
Certain (t)	$\bar{t}$	σ_t^2
Uncertain (T)	$\bar{T}$	σ_T^2

$$\bar{T} = \bar{t} + P\Delta \qquad (1)$$

$$\sigma_T^2 = \sigma_t^2 + P\sigma_\delta^2 + \left[\frac{1-P}{P}\right](\bar{T}-\bar{t})^2 \qquad (2)$$

σ = Time increment added to t by uncertainty
P = Probability that uncertainty will add δ to t
δ = Distribution: mean = Δ, variance = σ_δ^2

From (1) and (2),

$$\sigma_\delta^2 = K\Delta - \Delta^2$$

With

$$K = \frac{\sigma_T^2 - \sigma_t^2}{\bar{T} - \bar{t}} + (\bar{T} - \bar{t})$$

FIGURE 2.—An outline of the derivation of the coefficient K, the effect upon discrimination reaction time of channel uncertainty.

when he is uncertain, and these are denoted by the symbols in figure 2.

Let δ be an increment of time that is added to reaction time as a result of uncertainty. We need not specify anything more about δ, not even that some value of it is added by uncertainty on every trial. The parameter P is the probability that uncertainty will add such an increment on a particular trial. There is a distribution of values of δ, which remains unspecified, except to say that it has a mean that is called Δ and a variance for which the symbol is shown in figure 2.

If uncertainty adds some unspecified increment of time to reaction time on some proportion of trials, then the mean reaction time under uncertainty will be related to the mean reaction time under certainty in the manner shown by equation (1) in figure 2. Reaction times will, on the average, be lengthened in proportion to the mean value of the added increments and in proportion to the percentage of trials on which such increments are added. Similarly, the variance of reaction times under uncertainty will be related to the variance of reaction times under certainty in the manner shown by equation (2). This equation is just a general form of the familiar equation for the variance of a sum and assumes that the size of the added increment is independent of the value of reaction time on each trial. This assumption is exactly the assumption that the theory would make and equation (2) is compatible with the theory. Equations (1) and (2) can be combined to eliminate P with the result shown at the bottom of figure 2. The equation shows that the variance of the added increments is a function of a coefficient K and of the mean value of the δ distribution. The coefficient K can be calculated from experimental data as shown at the bottom of figure 2. It is a function of the extent to which uncertainty changes both the variance and the mean reaction time. The theoretical meaning of K is contained in the equation, which shows that K expresses a relationship between the variance and the mean of the distribution of the

increments added by uncertainty. The advantage of this is that K can be calculated from data in the manner shown, and that the value of K is independent of the proportion of trials on which uncertainty adds an increment to reaction time. Because of this it is not essential to control rigorously the direction of attention.

The Form of Discrimination Reaction Time Distributions

I would now like to describe a kind of theoretical model, which did not follow from the theory of attention, but arose as a result of a more detailed examination of some of the reaction-time data. This model not only seems to give us an additional interesting parameter, but it also provides an extension of the theory that is crucial for interpreting some of the parameters discussed in the preceding sections.

This model was obtained by plotting frequency distributions of discrimination reaction times for the certainty conditions and, at the beginning, simply abstracting common characteristics from the obtained distributions. Briefly, it was noticed that nearly all of the distributions had a span of approximately 150 milliseconds and that a very large number, well over half of them, were very similar in form. This abstracted form is shown in figure 3. The typical distribution has two peaks, one peak located about one-third of the way up from the lower limit of the distribution and the second peak, which is always a minor peak, is located about two-thirds of the way up. Furthermore, many of the distributions are very well described by three linear segments, as shown in figure 3.

An idealized distribution, like the one shown in figure 3, strongly implies a quantal process. Figure 4 shows a logical diagram of a message-transmission system that would generate a distribution of total transit times of the kind portrayed in figure 3. The quantity A represents the total fixed transmission time between stimulus and response; that is, the transmission time that is the same on different trials. The variance of reaction
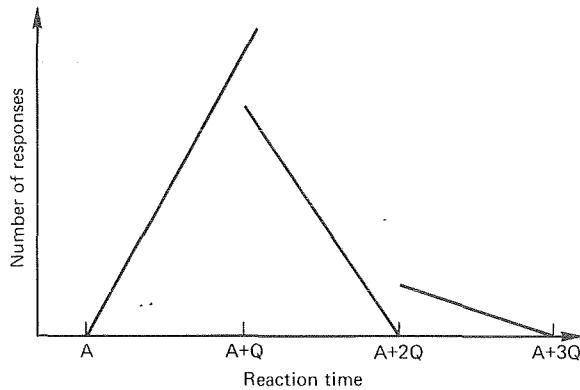
FIGURE 3.—A frequency distribution of reaction times generated by the model in figure 3.
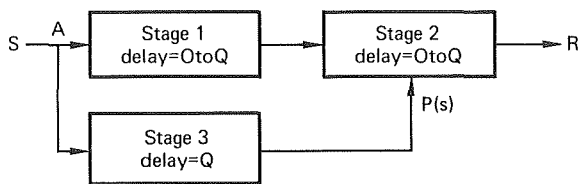
FIGURE 4.—A quantal model of information flow in discrimination reaction time.

time is contributed entirely by the three stages that are shown. A message enters stages 1 and 3 simultaneously and it is required to remain in stage 1 for some length of time that is equally likely to be any value from 0 to Q. It remains in stage 3 for exactly Q. After it leaves stage 1, it immediately enters stage 2, which has a delay characteristic the same as stage 1. If the message has left stage 2 before it leaves stage 3, then the stage 3 branch has no effect upon the response. However, if the message leaves stage 3 before it has left stage 2, then there is some probability $P(S)$ that stage 3 output will cause the message to be delayed one more Q.

I have no desire to defend this specific model and I present it mainly to indicate a direction in which I intend to do further work. This model is only an illustration of a class of possible models that have the important property of consisting of central delay stages having quantal characteristics. The one particular model shown here provides a satisfactory fit to a large number of obtained distributions, but there are many that are not well fitted by it.

Experimental Results

One comprehensive set of experiments has been completed in which the preceding measurements were all made on each of a number of experimental subjects. These data are available (refs. 3 and 4) and I will not repeat them here. Let me just summarize the major findings.

Enough data are obtained so that it is possible to estimate each parameter separately for each experimental subject. There is no averaging of raw data.

The major finding is that M, K, and Q are equal in absolute magnitude. The average value obtained for M was 54 milliseconds and K and Q both yielded means of 53 milliseconds.

The parameter K, which is determined from the effect of uncertainty upon reaction time, is determined twice for each subject, once for the visual signal and again for the auditory signal. These two values of K do not differ. The size of K is independent of sensory channel, and the value that is obtained for one channel is significantly correlated with the value obtained for the other channel over individuals. This same pattern of results is also obtained for Q. The time constant governing central processing, which is inferred from visual reaction times, is the same as that for auditory reaction times, even though average reaction times differ by a substantial amount.

I have concluded that these three major parameters directly reflect the action of a single temporal process, and I refer to this process by the term "time quantum."

Several years ago, after many years of wrestling with the idea, it no longer seemed reasonable to expect that the temporal characteristics of the electroencephalogram would bear any simple relationship to behavior. As most of you know, this was a widely proposed notion a dozen years ago, but it did not seem to lead to any convincing experimental evidence. With the behavioral

data that I have just described in hand, however, it was difficult to resist taking the next step, so we proceeded to obtain electro-encephalographic recordings from the same subjects. We did this for the obvious reason that the behavioral quantum appeared to be very nearly equal to one-half of the period of the alpha rhythm; that is, to the interval between zero crossings of alpha.

The frequency characteristics of the alpha rhythm are very well known, and, of course, it was no surprise to find that the interval between zero crossings agreed rather well with the measurements of the behavioral quantum. The purpose of the measurement was to determine whether there is a correlation between the magnitude of the behavioral quantum and the alpha zero-crossing interval. The data showed the existence of such a correlation.

There was, however, one discrepancy in these data which, although small, was consistent and could not be overlooked. I want to emphasize it here. Each of the three behavioral quantum estimates averaged about 53 msec. The average value of the alpha interval was consistently 48 msec. The difference is small but crucial. Furthermore, the distribution of values of the behavioral quantum was skewed in the upward direction, suggesting that there might be measurement error, at least for some subjects.

Theoretical Interpretation

These results suggest that the quantum represents a limit that is imposed on neural information processing. I have already hinted at an interpretation of this, which I will go into in greater detail now.

The linear form of the successiveness-discrimination function, and the conclusion that two independent sensory events must be separated by one quantum in order to be discriminated as successive 100 percent of the time, can be accounted for in the following way. To discriminate an event B as occurring at a later time than an event A when the two events cannot be attended to simultaneously, it may be necessary for the subject to note first the occurrence of

event A, and then switch attention to monitor event B. If event B is seen to occur following the switching of attention, then B can be judged as later than A and the two events can be discriminated as successive. However, if A is seen to occur and B is found to have already occurred when attention is switched to its channel, then the two events are effectively simultaneous.

If attention is restricted by the quantum in such a way that it can switch from one channel to another not more than once during a quantum, say at the end of a quantum, then the probability of discriminating A and B as successive will be equal to the probability that the end of a quantum falls in the time interval between the two signals. Because the time of occurrence of a signal bears no relationship to the time base of the quantum, the probability of a switching point falling between two signals will be simply the ratio of the time interval between the signals minus x and the duration of a quantum.

This leads to the expectation that the successiveness discrimination function will be a linear, one-quantum function like the one shown in figure 1. It is a corollary of this that the time that must elapse after a signal is presented before the next switching point occurs will be a duration between zero and one quantum and that all durations within that interval will occur with equal probability.

What is implied by the finding that the coefficient K is also numerically equal to one quantum? The most direct interpretation of this can be seen by looking at the theoretical equation at the bottom of figure 2. If the increment of time that may be added to reaction time by channel uncertainty is always exactly one quantum, then the variance of the increments will be zero and K will be equal to the increment itself, and, obviously, the increment itself will be its own mean. Therefore, the conclusion that uncertainty adds exactly one quantum of time to reaction time on some proportion of the trials and exactly nothing on the remainder of the trials is a sufficient conclu-

sion. If the added increment is due to the need to switch attention on those trials, then it follows that the time that is added to reaction time as a result of having to switch attention is always one quantum. This conclusion seems to be in conflict with the conclusion reached above in the case of successiveness discrimination. Yet, it is not in conflict because the time that is added to reaction time may not be the time required to switch attention.

This is where the third parameter is important theoretically. If stage 1 in the model shown in figure 4 has the characteristics ascribed to it, then, on a reaction-time trial when attention is directed at the wrong channel, the time that would have been spent in stage 1 is, instead, absorbed by the switching of attention and the message then enters stage 1 exactly at the beginning of a quantum and must reside therein for one full quantum. From this it follows that the time that is added to the total transit time for the message by the need to switch attention is always exactly one quantum. The time that must elapse before the next opportunity to switch will be, with equal probability, any value between zero and one quantum.

This synthesis is based on the assumption that the time base that controls the switching of attention is the same as that which determines how long the message remains in stage 1. This is consistent with the idea that there is a unitary quantum generator, and yet I must point out that it becomes arbitrary to an extent when we consider that stages 1 and 2 of the model in figure 4 are assumed to be independent of each other. The same time base controls attention and stage 1, but it seems that an independent time base with the same period controls stage 2.

The theory that is developing out of this work is one that deals with the temporal microstructure of human information processing. The concept of attention plays a central part in it. In many ways, the theory is similar to the filter theory that was proposed by Broadbent (ref. 5). The similari-

ties will be evident in the discussion of the structure of the theory, which will be presented below. There are also marked differences between the two theories. For example, the present theory has not yet found it necessary to postulate a short-term storage immediately prior to the attention mechanism. The reason for this is that the experiments that have been done so far have been designed so that they do not depend on the characteristics of such a memory unit. A short-term memory is irrelevant to our present purposes. The theories also differ in their degree of explicitness, the present theory being completely explicit within well-defined boundaries, to the point of permitting quantitative tests. The theories also differ in the kind of data to which each is thought to be relevant. Broadbent's theory, as is well known, was designed to account for the handling of complex verbal messages and other kinds of temporally extended behavior such as vigilance. The present theory, on the other hand, deals with elementary sensory signals and with events that occur over small fractions of a second.

A logical structure lurks behind the concepts that I have discussed above. This structure interrelates the concepts and I would like to discuss it next. It is shown diagrammatically in figure 5.

The Structural Theory

It is convenient to begin by defining a central processor through which some messages are transmitted between the stimulus inputs and the response outputs of the organism. There is no need to speculate about the operations that are performed by the central processor. It is sufficient to say that some messages pass through it and that appreciable lengths of time are required for their transmission. We go one step farther and suggest that within the central processor a message is transmitted through a series of stages and that each of those stages consumes a certain amount of time. The configuration of stages is highly flexible. Different configurations are formed to meet different requirements. The con-
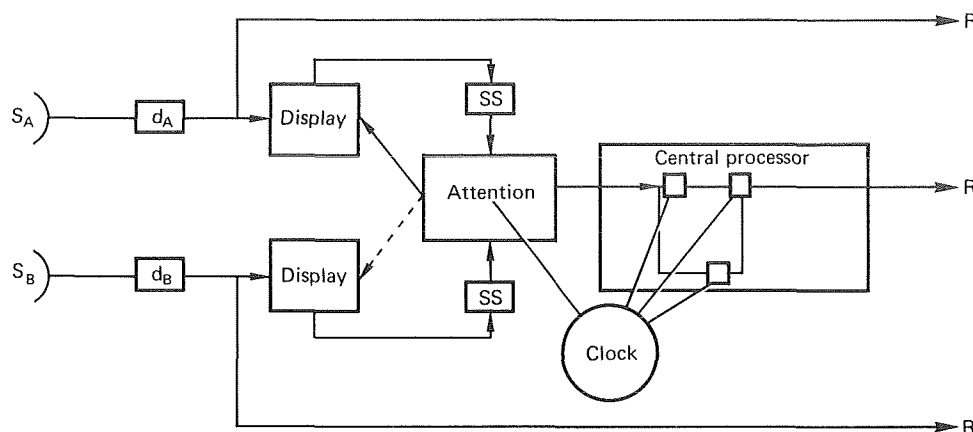
FIGURE 5.—Diagram showing some of the relations among the major concepts in the structural theory. S_A = signal presented to channel A; d_A = transmission time from receptor to display area; SS: switching signal.

figuration of stages is different for different information processing tasks, and, for a single task, different individuals may arrive at different configurations.

The admission of messages to the central processor is controlled by an attention mechanism. This mechanism is assumed to act as an all-or-none gate. Messages from only one sensory channel are admitted to the central processor at any one time. If attention is directed at a particular sensory channel at the moment a message arrives in that channel, then the message can be transmitted onward into the central processor without delay. But if attention is directed at channel B when a message arrives in channel A, the message is delayed by at least the time required for attention to switch from one channel to the other. One wonders what the factors are that determine which channel will be attended to at a particular time; there are undoubtedly many different factors that are influential in this respect, some of them to be found in the stimulus and others in various internal states of the organism. This theory does not attempt to classify or deal with these factors at the present time. However, it is important to point out that the theory places no restrictions on the way in which attention switches from channel to channel. No regular order of switching is assumed and

no particular channel is assumed to have priority.

Not all of the messages that are transmitted by the organism pass through the central processor. Some messages may be transmitted via bypass routes. One reason for saying this is the finding that reaction times that depend only on the detection of the occurrence of the signal, rather than on a discrimination between the signals, seem to follow different principles. Detection reaction times are different in two important respects. In the first place, channel uncertainty has no effect upon detection reaction time when the subject is highly practiced, implying that attention is irrelevant. In the second place, the variance of detection reaction times is smaller than the variance that would be generated by any configuration of states within the central processor of the kind being defined here. It is one of the long-range goals of this research to define more precisely these two classes of messages—those which do and those which do not involve the central processor. The further development of the theory may make that possible. That there are bypass routes must be recognized. Their existence certainly complicates the experimental program.

It is necessary to assume that the attention mechanism receives identifying signals from each of the many sensory channels.

When a message arrives in a sensory channel to which attention is not directed, the channel is capable of signaling the attention mechanism to switch over to it. This implies a partial interpretation of incoming information at a level prior to the processor and prior to attention as well. The degree of detail that is required in the analysis of information at this level cannot be known until we have an adequate definition of the concept of sensory channel and something approaching a catalog of sensory channels. The effects of switching signals are undoubtedly highly probabilistic and, as a result, the switching of attention from channel to channel is usually an unreliable and noisy process.

Incoming sensory messages traverse sensory channels. These channels begin at the receptor organ and end in a hypothetical information-display area. Time elapses between the moment of arrival of the stimulus at the receptor and the moment of arrival of the resulting message in the display area. This time, which is referred to as an afferent conduction time, is an important quantity. Because we do not know where any display area is located, it is not possible to identify afferent conduction time with any particular electrophysiological latency. It is necessary that these temporal variables be included as parameters in the psychophysical theory.

It is not reasonable to assume that a particular stimulus, no matter how simple, produces a message in only one sensory channel. On the contrary, I suspect that every stimulus produces many messages in different channels, and that these different channels have somewhat different conduction times. Most tasks probably can be performed by utilizing information in any one of the many channels in which the stimulus produces an effect. This implies that afferent conduction time may not be a fixed quantity, even under fixed stimulus conditions. If a subject selects his sensory information from one channel at one time and from another of the possible channels at another time, then the afferent conduction time may change. We have seen several different instances in our experiments in which this seems to be happening.

This is another very important reason why afferent conduction time must appear as a parameter and must be calculable for each particular kind of experiment.

A sensory channel, then, consists of a receptor, a transmission pathway, and a display area. Messages are admitted to the central processor by the attention mechanism from the display area of a sensory channel. The display area of a sensory channel is defined by its relationship to the central processor. A sensory channel consists of a set of all possible messages that can be admitted simultaneously to the central processor. In other words, a channel consists of all possible messages to which attention can be directed simultaneously. We must admit that we know very little about the organization of sensory channels that are defined in this way. We cannot state where the boundaries of the sensory channels are. Another long-range goal of this research is to provide the means for discovering the functional anatomy of the sensory systems that is implied by this definition. I have started by assuming that signals in different sense modalities, such as visual and auditory, generate messages in independent sensory channels. This does not, however, mean that I wish to identify sensory channel with modality; quite the contrary, I believe that there are also multiple sensory channels within each sense modality. We need a theory with quantitative power to enable us to map out these channels.

There is one more major concept to be discussed to complete this presentation of the theory, and that is the concept of a timing mechanism that controls the flow of messages in this system. The theory is specific about this and proposes that there is a timing mechanism, or a "clock," that controls both the attention mechanism and the central processor. The clock operates by generating a succession of time points. The evidence to date leaves it sufficient to state that under normal conditions these time points are generated at a fixed rate, a rate that is constant for a single individual but that differs to a small, although significant

and measurable, extent for different individuals.

The time points are generated at an average rate of approximately 20 points per second, which means that the interval between successive points is about 50 msec. This interval between successive clock pulses is referred to as the "psychophysical time quantum."

Two major functions are ascribed to the clock pulses at the present time. They control times when attention can switch from one sensory channel to another, and they control the flow of messages through the central processor by determining when the message can be transmitted from one stage within the processor to the next stage. These are two of the ways in which the clock coordinates and integrates the time flow of information.

What does this concept of a clock imply neurophysiologically? It suggests that there is some periodic process in the brain that is likely to be distributed widely throughout the brain and that controls the flow of information. It is known that there are extensive periodic brain processes. The most salient of these is the alpha rhythm of the electroencephalogram (EEG). In the human being, the alpha rhythm is a sinusoidal variation that has a very constant period. The voltage level crosses the zero axis approximately 20 times per second. It is possible that the alpha rhythm is one manifestation of the hypothetical clock. If this is the case, then we have an independent method for measuring the duration of the quantum, a method that is precise and very easy to apply. I do not want to imply that the electrical changes that comprise the alpha rhythm are involved in the information processing system in any causal way. For several reasons, that seems to be a questionable assumption and for the present time it is best to refer to the alpha rhythm as no more than a manifestation of the clock.

SOME ADDITIONAL EXPERIMENTS

There are three additional experiments that I would like to describe briefly. The first two of these have modified and slightly complicated the ideas concerning the successiveness-discrimination function. The revised definition seems to be a more satisfactory one. These experiments also confirmed the correlation between the behavioral quantum and the alpha half cycle. These experiments have been described in a recent paper (ref. 6).

One- and Two-Quantum Successiveness Functions

In the early experiments on successiveness discrimination, we did everything we could think of to make it possible for the subjects to perform at a maximum level and to encourage them to do so. This included giving them feedback on each trial, making their overall results known at the end of each session, and adjusting the difficulty of the discrimination so that it was a challenge to them.

Twenty-six subjects were run under such conditions. This usually involved from 30 to 40 experimental sessions for each subject. There is an initial practice effect, and a number of practice sessions must be conducted to bring performance to a stable level.

During practice, two of these subjects displayed an unusual phenomenon, which can be referred to as a quantal shift in performance. Very early in practice, both of them showed stable performance; that is, no systematic increase or decrease in total performance from session to session, over a period of 12 sessions. They were then changed to other experimental conditions that involved a similar task. Later, they were returned to the initial conditions and given additional training. In these later sessions, they reached a stable level of performance that was substantially above their initial level. Successiveness functions were fitted to their data for the early and the later sessions separately and it was found that in both cases the span of the function was reduced by a factor of 2 in the later sessions.

This change is shown schematically in figure 6 (a). The single linear segment spanning two quanta represents the subjects' performance during the early sessions, while

(a)

(b)
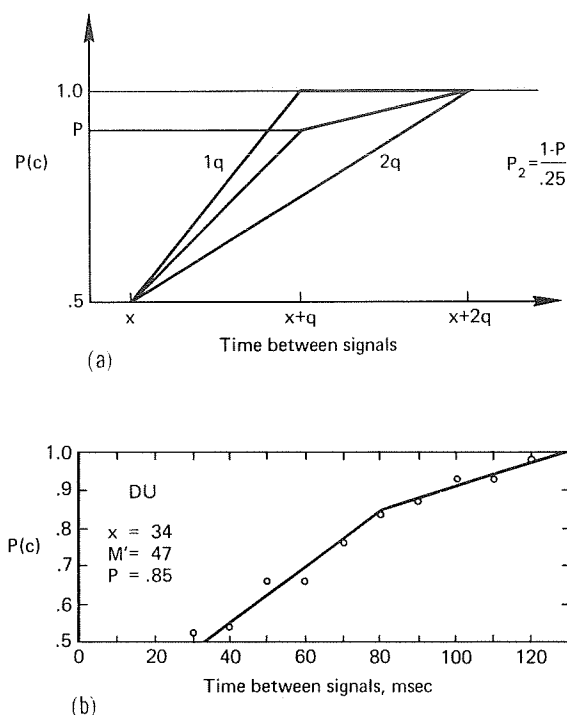
FIGURE 6.—(a) Theoretical one- and two-quantum successiveness-discrimination functions and their combination in the two-state model. $P_2 =$ probability of being in state 2; $q =$ quantum size. (b) An illustrative set of data for a single subject with the parameters of and the lines representing the best-fitting two-state function.

the one quantum line represents their performance after they had received additional practice.

This observation suggested the possibility that a subject may be in either one of two distinct states over a long period of time. In state 1, the span of the function is one quantum, and in state 2 it is two quanta. Which state the subject enters might be influenced by practice, motivation, or other general psychological conditions.

I decided next to determine whether the proportion of subjects who enter state 2 could be increased by manipulating feedback to the subject and task difficulty. Five new subjects were chosen for this experiment and they were run with the usual procedure, except that for the first 12 sessions no feedback of any kind was given

and the difficulty of the task was set at a relatively easy level.

Four of these subjects entered state 2 during these early sessions. The fifth subject went directly into state 1. Following the 12th session, feedback was introduced and the level of difficulty of the task was changed to the usual more difficult level. Additional practice sessions were conducted until the four subjects leveled off at a new, higher level.

The hypothesis that M is two quanta in state 2 and one quantum in state 1 was supported in this experiment, both by the mean values of M and by the high correlation between the two sets of M values. The evidence for the existence of the two states seems fairly clear. We have not continued to pursue the interesting problem of identifying the variables that determine which state the subject enters.

Two-State Successiveness Functions

Accepting the hypothesis of two states and the finding that the same individual may be in one state at one time and in the other state at another time, it is difficult to assume that all individuals can maintain either state over a length of time to the complete exclusion of the other state. Yet, that assumption is implicit in the earlier methods of interpreting successiveness data. A more realistic assumption would admit the possibility of a subject being in state 2 on some proportion of trials, even under experimental conditions that are designed to maximize performance. This might account for the earlier finding that the span of the successiveness function is slightly larger than the alpha half cycle.

This leads to a two-state model of successiveness discrimination in which the successiveness function is the weighted mean of two linear functions having the same value of x but spanning one quantum in one case and two in the other. The weighting factor is P_2, the probability of being in state 2. An example of this two-state function is also shown in figure 6(a). It consists of two linear segments, each of which spans

one quantum on the time axis. The point of intersection of the two segments occurs at a value of $P(c)$ that is related to the probability of being in state 2 in the manner shown in the figure.

Data that are adequate for analysis by means of the two-state model have been obtained for 13 new subjects. Electroencephalograms were also recorded for these subjects in each experimental session. As in the earlier study, the average value of the alpha half cycle was found to be 48 msec. The average value of the behavioral quanta determined by means of the two-state model was also 48 msec. The correlation between the behavioral quantum and the alpha interval was confirmed.

To obtain data that are an adequate test of the two-state model requires setting the level of difficulty of the task so that it is relatively easy; that condition was used in this experiment. Under that condition, it is interesting to note that the probability of being in state 2 was not close to zero for any subject. This means that the one-quantum successiveness function is approximately correct, but only under very special experimental conditions. The generality of the one-quantum function is quite limited.

Experimental Alteration of Quantum Magnitude

I should hasten to point out that we have not yet accomplished that which is implied by the title of this section. This is a project that is in progress. We have been working on it for the past 2 years and most of this effort has been borne by John Santa Barbara.

There are two main reasons why I want to find out where alpha fits into this story. In the first place, I am curious about the alpha rhythm itself. It is perhaps the most salient aspect of brain-wave recordings and we have never been able to identify its psychological or behavioral significance. In the second place, the temporal characteristics of the alpha rhythm can be determined easily, rapidly, and with a high degree of precision. With the proper equipment, one can accomplish this for a single individual in just a few minutes. This is in marked

contrast with the behavioral methods that we use; such methods require experimental sessions spread out over many weeks to complete a single measurement. Therefore, if it can be established that the conclusion about alpha that I am suggesting is a valid one, we would have a very powerful tool to use both for further analytical work and for any attempts that might be made to apply this theory in practical situations.

We need further evidence of a different logical nature concerning the relationship between alpha and the quantum. Most of our evidence is at the present time only correlational and it is not adequate to support a strong conclusion.

Another approach is to try to exert experimental control over the magnitude of the quantum. If we could find a way to change the frequency of the alpha rhythm, we could then make measurements to see whether we get the same changes in the behavioral quantum. This project is an attempt to do that. Mr. Santa Barbara set out to try to find some agent or combination of agents that is effective in slowing down or speeding up the human alpha rhythm. This strategy was taken because of the greater ease of measuring the EEG, and it is more efficient first to find an agent that produces a reliable effect upon the EEG.

We hope to find an agent that will shift the frequency spectrum in the alpha range and leave the remainder of the spectrum unaffected. Many agents have been tried, most of them selected on the basis of reports in the literature. They include dextrose, fasting, oxygen inhalation, carbon dioxide inhalation, Diamox, alcohol, Librium, and Dexedrine. None of these agents produced consistent results. Most of them had very little effect at all, and when effects were noted they were in one direction for some subjects and in the opposite direction for others.

We would like to produce a change in peak frequency of at least 20 percent. In no case have we been able to produce a change that large. With some subjects, some agents appeared to produce a change as large as 10

percent, but even these changes, the largest ones we obtained, were very inconsistent from subject to subject. Speaking statistically, the results that we have obtained are consistent with the results recorded in the literature. Agents that are reputed to speed up the alpha rhythm do tend to do that, and those that are reputed to slow it down tend to do that. The effects, however, are inconsistent and very small in magnitude.

On the basis of the work we have done so far, we are led to conclude that the alpha rhythm is highly impervious to external modification and highly stable for a single individual over long periods of time.

We are continuing the search for an effective agent. We are currently investigating the influence of body temperature. There are many reports in the literature concerning this. These studies have reported consistent effects and have often shown in detail the functional relationship between alpha frequency and body temperature. We also get consistent effects with body-temperature changes. Once again, however, the changes do not exceed about 10 percent, at least over the temperature range that we have used.

I would like to mention an experimental result that seems to be emerging, but is still quite tentative. We have been trying to develop a method for changing body temperature that causes as little discomfort to the subject as possible. In the course of doing this, we are finding that if you change body temperature using a procedure that depletes water and salt, then you get the usual change in the alpha frequency. However, if you take steps to prevent the depletion of electrolytes and fluid, then, even though body temperature changes just as much, alpha frequency does not change.

FUTURE RESEARCH

My research during the next 2 years will be directed mainly at two tasks: a more intensive experimental and theoretical analysis of reaction-time distributions and the influence of channel uncertainty; and a continuation of the attempt to alter quantum size and to apply this to a study of the relationships among the major parameters.

It required most of the past year to arrive at a more adequate definition of the one parameter M. The next step is to study reaction time in the hope of reducing measurement errors in those experiments. Also, there are other experiments underway that are designed to extend the theory to encompass other phenomena; these will continue.

This work promises to have generality because it seems to integrate very different kinds of data into what is still a simple and coherent theory. However, all the work has been done in only one laboratory and even the basic experimental results are not yet firmly established. It is my conviction that the generality of the theory is sharply limited at present. In 2 years, at the present rate of progress, I may feel that the basic results are established. We will then be in a position to test the generality of the theory along other dimensions.

Useful empirical generalizations will be difficult to acquire. For example, consider the question of successiveness discrimination. I commented above that the experiments on the two-state successiveness function demonstrate that the one-quantum function is useful only under very limited conditions. The two-state function introduces a new parameter that has been shown to influence this behavior extensively. I am afraid that I can only make crude guesses about the variables that influence this parameter. Furthermore, I know from a number of other experiments that successiveness functions at the present time have meaning only within a very restricted discrimination context. They have all been obtained using a forced-choice method. Other psychophysical methods, which seem more similar to situations that one might encounter in applied settings, give results that I cannot reconcile with the forced-choice results at the present time. In a yes-no context, there seem to be still other parameters operating, and we have not yet succeeded even in identifying them.

As another example, take the conclusion stated above that channel uncertainty adds a quantum to reaction time on some trials.

For one thing, I know that this conclusion applies to a situation that involves a single response. Whether it would apply if the subject also had to select among responses is something that we simply do not know. I am quite sure that we would find a much more complicated set of relationships with multiple responses. Also, I know that this conclusion does not apply to detection reaction times. Finally, I also know that if one complicates the stimulus configuration in the discrimination reaction time experiment by doing nothing more than adding a fourth signal, an additional negative signal in the auditory channel, the conclusion does not apply. Distributions of the parameter K that are then obtained do have a peak in the vicinity of one quantum, but they have a second peak at the two-quantum level and the average value of K is quite different than it is in the three-signal case.

We also do not know anything about the possible effects of stimulus parameters, such as intensity, on the measured extent of the quantum, nor do we know whether similar results are obtained for combinations of sensory channels other than the one that has been used in all of my experiments.

A NOTE ON APPLICATION

Despite the paucity of precise generalizations that would be useful in solving applied problems, I believe that it might be helpful to use this theory, or parts of it, in formulating certain applied problems and in designing relevant experiments. I can conceive of this being done in several different areas, assuming that information-processing efficiency is an important issue, such as designing displays, selecting personnel, monitoring the state of an operator, or evaluating the effects of environmental conditions, including physiological conditions.

The theory contains at least three general classes of parameters. There are four parameters that converge upon the quantum concept. Other parameters have to do with message transmission times, mainly within sensory channels. And, yet, other parameters are probability values, such as the probability of being in a particular configuration or information-processing state, and the probability of attending to a particular sensory channel.

From my point of view, the quantum concept is the most important of these because it is the source of integration for the theory, and I am sure that I have communicated that attitude in this presentation. However, the priorities that I attach to the various parameters are not necessarily those that should be assigned by someone who wished to use the theory in other contexts.

The magnitude of the quantum, except insofar as it establishes limits upon performance, is probably of minor significance as far as raw performance is concerned. I say this because individuals differ so little in quantum size, minimizing the value of selecting individuals according to this criterion, and also because it seems unlikely that the size of the quantum can be altered to any important extent.

On the other hand, quantum size may become important in complex tasks that involve many serial stages of information processing. Performance variance in complex tasks may be determined strongly by quantum size. However, we know very little at present about the extent to which the quantum concept is relevant to complex tasks.

I can illustrate this point with some of my own data taken from the two-state successiveness experiment (ref. 6). If one forms two groups of subjects by taking the six with fastest alpha and the six with slowest alpha, one can then compare the two groups with respect to their overall performance on the successiveness task.

The correlation between alpha and the behavioral quantum is close enough so that the average values of alpha and q are within $\frac{1}{2}$ msec of each other for both the fast-alpha and the slow-alpha groups. In spite of this, the groups differ hardly at all in overall performance and, in fact, the slow-alpha group shows slightly better overall performance. An inspection of the other

parameters clarifies this result. The groups are identical with respect to x, but quite different in the probability of being in state 1. The group with the larger quantum size was much more likely to be in state 1— enough to more than compensate for quantum size ($P_1=0.57$ for large q group and 0.25 for the other group).

The implication of this is, of course, that some of the parameters other than quantum size may be much more potent determiners of performance differences.

DISCUSSION

LLOYD A. JEFFRESS: Have you tried getting your stimuli in accordance with the phase of the alpha rhythm?

ALFRED B. KRISTOFFERSON: No; I have not tried that yet. I am reluctant to try it because there would be variability in the conduction pathway that would be hard to take into account.

HAROLD G. MILLER: What about electric shock as a method of changing your timings?

KRISTOFFERSON: Well, I have not tried that, but it would only be a very momentary thing.

JEFFRESS: Another possibility would seem to be to use a different sensory modality, say, touch and hearing, and use vision to drive the alpha rhythm.

JOHN W. SENDERS: Yes; possibly you could drive the alpha with touch.

JEROME I. ELKIND: Could you clarify one thing for me? It seems to me that you are talking about a clock running at 50 milliseconds while at other times you are talking about a delay of 50 milliseconds. Are you always assuming you are in synchronism with the clock?

KRISTOFFERSON: No; never. I am always assuming that the stimulus is completely independent.

ELKIND: Yet you said that things are going to be held there in stage 1 for 50 milliseconds.

KRISTOFFERSON: Because stage 1 and attention are controlled by the same clock; so if the message is held out of stage 1 waiting for attention to switch, then when it is admitted to stage 1, it will be admitted at the beginning of the quantum.

JOSEPH MARKOWITZ: It is synchronous gating.

MILLER: You only tried two of all your five senses.

Maybe you could try heat too, which would be the third.

WARD EDWARDS: The reaction-time questions you are dealing with are very similar to those dealt with by people interested in the tradeoff between speed and accuracy. You did not say much about accuracy. I assume it was very high.

KRISTOFFERSON: It was. The probability of a wrong response was below 5 percent, between 2 and 5 percent.

EDWARDS: Suppose you manipulate the situation by means of your payoff matrix in such a way that you get 20 or 30 percent errors. You will expect to get substantial decreases in reaction times. How would that fit into this quantum conception?

KRISTOFFERSON: Well, I do not know what the effect would be on K; that is not immediately obvious to me. Parameter K is a relation between two experimental conditions, both of which would be affected in the manner you are describing. Whether there would be a net effect on K itself, I cannot say. I do not think there would be. Now, the other one, Q, I do not know. I would be sure that would be badly damaged.

MARKOWITZ: Do you have any thoughts on how you can go back from an NQ state to an $(N+1)Q$ state? You were able to show a switch one way toward N. Then, do you have any feel for how you might go about getting them back?

KRISTOFFERSON: They would not go from the one all the way back to the two but they would go in that direction. In other words, P_2 would increase, but never all the way to one, once they had been in the one-quantum state.

REFERENCES

1. WHITE, C. T.: Temporal Numerosity and the Psychological Unit of Duration. Psychol. Monographs, vol. 77, no. 12, 1963, pp. 1–37.
2. HARTER, M. R.: Excitability Cycles and Cortical Scanning: A Review of Two Hypotheses of Central Intermittency in Perception. Psychol. Bull., vol. 68, 1967, pp. 47–58.
3. KRISTOFFERSON, A. B.: A Time Constant Involved in Attention and Neural Information Processing. NASA CR–427, 1966. (Available from Clearinghouse for Federal Scientific and Technical Information, Springfield, Va. 22151.)
4. KRISTOFFERSON, A. B.: Attention and Psychophysical Time. Acta Psychol., vol. 27, 1967, pp. 93–100.
5. BROADBENT, D. E.: Perception and Communication. Pergamon Press, Inc., 1958.
6. KRISTOFFERSON, A. B.: Successiveness Discrimination as a Two-State, Quantal Process. Sci., vol. 158, no. 3806, 1967, pp. 1337–1339.

Comments[1]

JOHN W. SENDERS

Bolt, Beranek & Newman, Inc.

JOHN W. SENDERS: My main purpose is to act as discussant of the presentations of Tanner and Kristofferson. My concern is not with the quality of the research that either of these scientists has carried on. Rather, I am concerned with the relation of this research (and its extensions and interpretations) to some of the problems that NASA faces in operational missions and in the real task requirements.

Psychologists do not invent mathematics to fit psychology. They very often invent psychology to fit mathematics. I have been impressed with the marvelous ways in which this goes on. The particular mathematics into which behavior is being fitted—this procrustean bed—is signal detection theory in the one case and a sort of quantum mechanics in the other case. If Elkind were here, it would be linear-servo analysis. I use information theory, and so forth. There is nothing terrible about this, but one has to remember that the mathematics is a convenience that, in many cases, circumscribes what one says about what really happened. The statements that were made earlier, about the necessity for face validity if one is to extrapolate from laboratory work (particularly in simulations) to operational situations, still hold.

Tanner's presentation elicits the following thoughts. In the laboratory, we study the detection of signals in noise, and the identification of signals that differ in small degrees from some other signal. We vary the probability of occurrence of two signals and ask the subjects to make estimates and to emit responses. From the data we calculate certain numbers that go into a particular mathematical model. The model leads to further inference about hypothesized internal variables in the operator.

An alternative kind of research is that in which people are put in simulators or real aircraft and allowed to fly. We have been very well informed about the different kinds of situations that can exist and the different kinds of decisions that are made. However, for certain critical situations, the number of cases that can actually be observed is small, and the opportunity to get information from the people who end up there is even smaller. As a result, we lose that information most useful to us.

Yet, there is an in-between type of study. Tanner suggested that an experimenter could use a television screen to present pictorial situations that might be characteristic of those that a pilot sees on breaking out of the clouds. The experimenter might then ask the pilot to make some kind of judgment as to whether he would go around or land. Presumably, these situations could be scaled so that one could know whether they were, in fact, good or bad. I think that this kind of experiment can closely approximate operational situations. At the same time, such experimentation can preserve the applica-

bility and utility of the signal-detection model.

A student at MIT is doing experiments on an automobile driving simulator that I think could be replicated in aircraft simulators. A model car is driven remotely with the help of television. The car bears a camera that looks at a model road. In general, the dynamics of the situation are, as nearly as possible, the same as those of a real automobile. The driver has a curving road to drive through with miniature traffic cones along the edge of it. At a particular point in the path, he is required to do one of two things, depending on the nature of the experiment. In one experiment, he merely estimates that he can or cannot go through that particular obstacle course without knocking over any traffic cones. Knocking over a traffic cone is analogous to a bad landing. In the other kind of experiment, he gives numerical estimates of the probability that he will, in fact, be able to go through the winding road without knocking over any cones. Then, of course, he continues driving and goes through the winding road. So you end up with the data that you need for a signal-detection analysis of the behavior. You have the estimated probabilities for situations that vary over a wide range. You also have actual data as to whether the driver did, in fact, knock over the cones.

I think this could be done in an aircraft simulator in the same way. At the moment that a pilot broke out of his television clouds, he could be required to make an estimate and then to proceed with the landing. In this way, an experimenter could accumulate statistics on the probability of successful landing following a prediction of success. Admittedly, this would be a very long and arduous program of research, but compared with "Alluisi's heroism," as Yntema has called it, it is relatively easy. Six months of devoted work would, perhaps, produce a sufficient body of data to permit a complete analysis.

Of course, one would also like to find out the extent to which one can use these results to make predictions about the behavior of

the same individual flying a simulator or a real aircraft. In other words, is there a consistency in performance across situational variables? Probably not, but the in-between type of study, in which the simulation devices that Rathert talked about would be combined with the signal-detection studies that Tanner talked about, might produce useful information.

Kristofferson's work elicits the following. What is the relevance to operation problems of fundamental research on hypothesized internal variables; the switching of attention or the switching of the on-line modality from, say, vision to audition or to touch or to olfaction or anything else? It has been claimed and, I hope, verified that there are well-defined, almost invariant, time quanta. Within these "moments," certain things happen that are necessary precursors to certain other things happening, and so forth. What is the relevance of this discovery? I think it is this: The discovery enables us to change our evaluation of systems from an a posteriori one to an a priori one. If we imagine some kind of a machine that is to go somewhere and do something, we can use the science of aerodynamics to calculate roughly what is going to happen; that is, what kind of signals are going to flow in the system. It is very difficult to calculate, a priori, the ease or the difficulty that a man is likely to encounter in operating the system, or to decide how much of the system must be automated or how one must allocate the workload between automatic devices and manually controlled devices.

On the other hand, assume that we could take the calculated information about the behavior of the system and, from this information, derive estimates of the loading to be placed on the human operator. At relatively low cost, we could then make a great many decisions that, at present, are almost inevitably preceded by the construction of simulators, prototypes, flying prototypes, training of pilots, and so forth. The components of switching, which Kristofferson reports, would constitute, in a sense, the elements of demand upon the operator.

From these components, we could conceivably calculate workload by examining the time sequences of events that might occur in a complex machine. (A great deal more work, of course, would be required.) However, if we had a completely predicted time sequence of events, we could calculate how many of such events the human operator could deal with in how much time. We could calculate the probabilities that for given amounts of time, certain events would not be observed, listened to, or looked at. In other words, by breaking a task down into a very large number of very small components, we might be able to predict the results of the sort that Alluisi gets on his performance-testing machine. I am glad that this is not dead; in a sense, it is what Kristofferson is doing. I think the possible success of his work will depend upon a more sophisticated approach to the problem with more a priori analysis than the simple assumption of linear additivity, which was the one held nearly 100 years ago.

Workloading should be one of the major criteria upon which decisions are made concerning whether a system is good or bad. If one can decide this beforehand, it is, of course, much better than deciding after the system is built. But what is wrong with the other methods of estimating workloading? For example, the measurement of system performance. A number of years ago, Frank Taylor demonstrated quite conclusively that system performance is not a good way of finding out whether a system is good or bad. The adaptive nature of the human operator of the system tends to wipe out many of the effects of variations in system parameters. If these effects exist at all, they are probably reflected in some internalized and probably unmeasurable change in the level of effort required for the operator to maintain system performance at a constant and desirable level.

Another way of measuring workload is that of the auxiliary task. Alluisi's machine is a multiplicity of auxiliary tasks, in fact, all auxiliary tasks. There is really no primary task. Yntema mentioned that the use of auxiliary tasks in recent years has made it possible to increase the sensitivity with which our testing devices react to minor changes in system characteristics.

The use of auxiliary tasks is usually based upon a view of the human being as a single-channel device. In my own research I have found certain problems with this view. Human beings are not quite a single channel. The components of the task are not additive, so you have to be very careful about how you put things together. You cannot independently measure a button-pushing task and then a mental arithmetic task and expect them to sum up. If you put two button-pushing tasks together, very often you get near additivity. However, when you start putting probability monitoring, button pushing, tracking, and a lot of other things all in the same pot, it is rather like mixing up oil and water. You may get somewhat less volume than the sum of the two volumes. There are, perhaps, more extreme examples, but my chemistry is very weak.

I would say that the approach suggested by Kristofferson is still gestating. We do not have a set of usable numbers with which we could go into a real system, even a very simple one, and make explicit predictions about the behavior of the people.

For a moment I am going to talk about my own work because it is very similar, although there are a few differences. I deal with explicit behavior, which is easier than dealing with hypothesized internalized variables. I think I have been doing this a little longer than Kristofferson and I have gotten some results that I can use. But I have been aware for many years that there are a great many switching functions going on. I have modeled my man, in a conceptual way, as a device with many sensory modalities. Some of these modalities, for example, the eyes, are able overtly to direct themselves at different points in space. Others, such as ears, in the case of humans, cannot be directed. (If we were donkeys, we might be able to do better on that.) It is difficult to tell where somebody is listening, but you can usually tell where somebody is looking.

Internally there are switching functions that may make a selection, as Kristofferson would have us do, between eyes and ears. I think even within a sense modality there are switching functions that select a particular attribute of either the visual or the auditory stimulus to which attention will at that particular moment be paid. If Kristofferson is going to give us useful information, I think he must eventually get down to that microscopic level.

I shall talk about my own work briefly and then come back to Kristofferson. I have been dealing with the questions: Where do people look? Why do they look there? In particular, I have been concerned with the intervals between observations of any particular stimulus or signal. For certain well-defined situations, the visual sampling behavior of the human observer can be very well predicted on the basis of a sampling model. Subjects behave in a lawful way; if mathematics says they must do something in a particular way, they do it that way. When people want to get information from a time-varying signal (rather like the one Alluisi is using in his probability monitoring), they have to look at the signal in a way that is directly related to its effective bandwidth. On the basis of some elaborations of this simple notion, I and my colleagues at Bolt, Beranek & Newman have constructed models and theories of how people are going to look at a world composed of a large number of visual signal generators. It turns out that people pretty much obey the laws as we lay them down so that one can predict correctly much about where they are going to look, how often they are going to look there, how long they are going to look, what the transition probabilities are going to be between instrument A and instrument B, and so forth.

The power of this approach is demonstrated by some recent work of Warren Clement, Henry Jex, and Dunstan Graham (of Systems Technology, Inc.) that combined the describing function work of McRuer et al., on the one hand, with the scanning work of Senders et al., on the other. They took the equations of the Boeing 707 and closed all the loops with their human operator to calculate the displayed signal characteristics. Then they applied a modified sampling model and calculated the frequency with which each of the displays would be observed. On the basis of the observation probabilities, they calculated the transition probabilities—it is a simple Markov process—between the instruments. On the basis of a very simple cost analysis, it was possible to lay out an instrument panel that would minimize visual scanning movements. The panel of 10 instruments mapped one-for-one onto the existing panel.

Either we wasted our time, because people obviously already know how to design instrument panels; or, alternatively, if we can postdict a system that has been the product of an "evolutionary" process operating over some 50 or 60 years of aviation, then we can probably predict what needs to be done in systems that are much different from present ones.

STEVEN E. BELSLEY: Another possibility, too, is that your data are so conditioned by the 707 panel that you cannot come out with anything else.

SENDERS: No. These are theoretical data.

BELSLEY: Where did you get them?

SENDERS: According to McRuer, they used the dynamics of the 707 and applied their mathematical men to them. There are no real pilots involved. All this is the composition of two mathematical models of a human operator with a real system and a prediction from that of how the panel would be laid out for this mathematical man. It turns out that this is how the panel is laid out. The people who design airplanes are not fools.

Because Kristofferson is also dealing with the times involved in the directing of attention and in the switching of attention from stimulus to stimulus or from channel to channel, his work may make it possible to compose good estimates of the demands that would be made on an operator by a system that has not yet been built but only hypothesized. I feel that this kind of fine-structured

analysis of behavior will, in the end, lead to a great deal of analytical power useful for the design of systems.

I will comment briefly on Alluisi's multiple-task performance battery. It is, of course, heroic. One thing that occasionally concerns me about that approach to performance testing is an assumption of additivity that is inherent in the composition of such task batteries. It is assumed that each task uses some well-defined fraction of the human being's capacity, and that these fractions can be added up. Then, when you are at 99.9 percent, it is very difficult to add something else.

I am also concerned by the lack of physical coherence in the system that the man is called upon to operate in such tests. The task is different from all physical systems that men control and operate. There is no set of necessary relationships, no behavior of an integrated system reflected to the man through a multiplicity of channels. The data are isolated rather than provide the operator a coherent picture of some external reality. I think that the behavior of people in aircraft is probably conditioned very much by such an integrated picture of the whole.

I remember, many years ago in our less-sophisticated days at Wright Field, that people would talk about instrument lags. The rate of climb indicator lags terribly. "You never know where you are. You always know what you were doing some time ago." Someone suggested that it might be better if we delayed all the instruments uniformly so the operator would know what the entire system is doing some seconds ago rather than giving him an oblique slice of the system in time. I am not sure that that is a good idea, but it is the kind of thing one thinks about when one thinks about coherent physical systems as opposed to a set of discrete stimuli which from time to time demand response and in which predictability is remarkably lacking from one signal to another.

EARL A. ALLUISI: I do not claim that we have an ideal battery. I think we have one of the best available today, but that does not mean that it should not be improved. In fact, we are continually doing research to develop new tasks. We are trying to get more of the functions represented in a way that will enable us to identify the measurement and talk about it.

We do not really make the assumption of additivity. I think the most important thing about the approach is something we have not talked about. I spoke earlier about the domain of work behavior. There are two domains: one of test behavior and one of work behavior. Man does not operate the same way in the two. I think we all recognize that psychological variables, such as the domain in which a man places himself, his motivation, and others, may contribute more to the variance that we are measuring than do some of the variables that we are controlling.

If I want to talk about performance assessment or about work behavior, my subjects must be in the appropriate domain—in the domain of work behavior, not in the domain of test behavior. Our failure to recognize this distinction in most of the work completed prior to World War II explains our failure to show decrements in performance with any of numerous stresses. I can make predictions today regarding the outcomes to be expected with certain different approaches. We have used some of the other approaches in our own studies. For example, in conducting the sleep-loss studies, we have used intelligence tests and we found that our subjects, after 40 hours of sleep loss, tested out the same as when they were fresh. However, a man after 40 hours of sleep loss is not able to operate at the same intellectual level as when he is fresh from sleep! You can observe that in his gross behavior. Our tests cannot, however, discriminate between the man with sleep loss and the man without. This is partially because the man places himself in different domains of behavior in the two cases. With the paper-and-pencil intelligence test, he places himself in the domain of test behavior. There he is remarkably able to call upon his performance reserve and to perform at the

same level as when unstressed. (Even in the test-behavior domain, you can show a decrement with sleep loss if you have a task that is highly dependent on attention for good performance.)

No, we do not really assume additivity, although it may have seemed that we were when we attempted to derive a general index of performance. The general index used was logically the best we could devise. What we need to have represented in the battery with a clear measure is something that I can identify and that you will agree is a measure of what I say it is—one of the different functions or processes that men are called upon to perform in different systems. I would hope to be able to determine which of these measures deteriorate with different stresses, and to what degree they deteriorate differentially. The problem of putting every-thing together, mixing it up, and predicting the real-life thing, is that it is dependent upon the development of a task taxonomy and task analysis that says how much of each function goes into each task, and what the interactions are. They are not going to be linear, I can predict that. So I agree with you fully—simple additivity is just too simple!

I wish also to comment on the supposed lack of physical coherence in the system. When I present the system and talk about it to the man working it, he regards his task as a coherent job. He accepts it. We have had over 100 subjects, all of whom have indicated their acceptance of the task battery and their perception of it as defining the job they have to do. The separate tasks are time shared and related, just like on any other job.

DISCUSSION

LLOYD A. JEFFRESS: It may be a coherent task from the standpoint of the subject, but the inputs have no relation to one another.

SENDERS: That is really what I meant. It does not have to be a mass with wings attached to it. It just does not present a picture of a meaningful world because the signals are completely independent.

WARD EDWARDS: Do you know whether that matters? I do not know of any research comparing coherent with incoherent worlds.

SENDERS: I do not know that it matters. My intuition tells me that it might matter quite a bit.

ALLUISI: There is a kind of coherence—the kind that comes with putting a job together and accepting it as a whole job. If we consider the job of piloting an aircraft and do not know much about it, it does not look like the elements in it form a coherent whole either. What I am saying is that there is a kind of coherence, perhaps because these people accept it as a job.

I want to discuss figure 16 in Alluisi's paper, "Pilot Performance: Research on the Assessment of Complex Human Performance," presented previously.

All experimental subjects became ill; not all of them showed decrements in performance. In fact, I found in this study the greatest differences among individuals that I have ever encountered as a psychologist. In this particular case I was enthusiastic—and still am—because there is an opportunity, which we do not have in other studies, of finding relations with the body chemistry. In this case, we know that performance is bound to deteriorate to a maximum degree while there are maximum body chemistry changes taking place during an illness. So we maximize the opportunity of finding correlations between performance and physiological variables simply by maximizing their ranges of covariation. Of course, there will be much to study afterwards to determine which of these correlations are real and which are spurious. We hope to do that.

What happens to performance? It seems to lag the illness in its changes. If you recall, temperature started up in the middle of the seventh day. Performance starts down on the eighth day and continues down a little after the point where treatment started (and temperature started down), then, later, performance picked up again. At the end of the study, the experimentals are still about 15 percent below the control subjects. They have not recovered yet. This second double drop on the eleventh day, which occurred in the two control subjects as well as the eight experimentals, has occurred in each of our studies that had a clearly defined period of stress. It occurred in both the sleep-loss and the illness studies. It is a demonstration that one of the psychological variables, motivation, is potentially accounting for at least as great a portion of the variation as anything else. The subjects have been sick, they were started on treatment, and they are now feeling pretty well again. As far as the subject is concerned, his job is done. He was sick, his performance fell; he got well, his performance rose. He asks, "What am I here for? Why do you not let me out?"

In the normal course of experimentation, I would have gone in just prior to the time that I expected this second drop in performance to tell the subjects,

in effect: "I want your motivation to continue at the same level in order to see how the recovery takes place." Unfortunately, I did not get in because these data were collected during the blizzard of 1966, and I was stuck in a motel, cooking for 300 people! The next morning the chief experimenter called me and said that performance was beginning to deteriorate. I said, "Well, you go in and give them the spiel." He did, and the performance began to rise again. The rise is due strictly to motivation in both the controls and experimentals. The fact that the experimentals are unable to get back up to the line of the controls shows that they have not recovered fully—not behaviorally, anyway.

Average performance did drop during illness; the drop averaged about 25 percent over the eight subjects. However, the individual differences were so great that we had one subject who showed essentially no decrement in performance while running a rectal temperature of 105° F, complaining of severe headache, and showing all the other symptoms of tularemia.

JOSEPH MARKOWITZ: What was his baseline performance in comparison with the others?

ALLUISI: We cannot find a correlation with it. He was neither better nor worse to start with. He was essentially average in performance.

SENDERS: How do you control for the baseline level? Is there any way you can guarantee that the people when generating baseline data are maximally motivated and working at peak performance?

ALLUISI: No.

SENDERS: Then his baseline performance might not have been at his peak; he might have reserve capacity, so he could maintain performance?

ALLUISI: No; I do not believe that is likely. Although we have individual differences in performance, subjects are still remarkably alike. It is, I suppose, something like playing a pinball machine. There are elements intrinsic to the task, which begin to make the people more alike. The battery itself begins to form behavior. Although persons may start off somewhat differently, they become more alike as they continue to work the task. This is, I believe, one of the reasons that knowledge of results is important to every phase. They are getting full knowledge of results all the way along. Where it is not intrinsic to the task, I present it explicitly so that the task is instructive.

To pursue the topic of individual differences, however, I had just reported to you that we had one subject who showed no decrement. On the other end of the scale, we had one subject who became nonresponsive to external stimuli. By that I mean that a physician could stick a pin in his leg, and he would not move the leg, or say anything. We had another man who had a 3 percent drop in performance, which is essentially no drop at all. We had men so ill that they had to be wheeled to their duty stations, but they came to work in order to perform their group tasks. A man would come on duty just because he did not want the crew to be penalized.

Panel Discussion[1]

STEVEN E. BELSLEY, *Chairman*
Ames Research Center

STEVEN E. BELSLEY: I gather from Swets that this is to be a panel of the whole group and anyone may participate. I would like to say one thing before we get started. I was amazed to find out that Senders was using what I call an engineering approach to certain problems. I thought he represented the psychological approach. To suggest that one would run through a representation of the real-life situation and then make a measurement and do this repetitiously sounds like an engineer's approach to the problem.

JOHN W. SENDERS: I am glad to join the company of engineers.

BELSLEY: Nowhere in this discussion (even Rathert did not bring up this point) has it been mentioned that the mechanism that has been used in gaining all those beautiful time histories, to evaluate the situation and to find out whether it is go or no-go is a device called the vocal controller, which is the pilot. We have tried to bypass the use of actual performance measures of the various processes that are going on, and have used instead an overall integrated feeling of how the performance goes. We have been successful in the past in using the pilot rating to measure how things are doing and how well the system works. Yet no one seems to want to correlate decisionmaking with

pilot ratings—if it is correlatable. Does anyone have any comments on this score?

DOUWE B. YNTEMA: Do you mean that pilot rating is a decision?

BELSLEY: You operate a system to do something. The question is, then, not necessarily how well the pilot does the job, but how hard he is working. Is he working at top performance? Halfway? Or is the task just plain easy? The pilot can give a vocal rating. Instead of finding out from someone else that he is working at 100 percent capacity, the pilot says, "I was working at 100 percent capacity."

The question remains as to how one predicts where he is going to come apart at the seams?

JOSEPH MARKOWITZ: Are you asking that we get the pilot to tell us after how many hours' sleep deprivation he will, in fact, begin to deteriorate?

BELSLEY: It has in the past proved useful. I just wanted to point out that nobody talked about it at all.

MARKOWITZ: It seems close to what Senders described when he said that you ask people to estimate the probability of success and then to do the task. That would be similar to asking a pilot, "How are you going to do after 4 days?"

BELSLEY: Or after he is there 4 days you say, "How do you feel? Do you think you can hack it?" Otherwise you might look at all your data and conclude that the experiment is going along really well.

WARD EDWARDS: There are some data on

[1] A general discussion of the papers presented and issues raised during the day. The discussion centered around measurement, prediction, simulation, and motivation of pilot performance.

the problem Senders raised, and bearing on the question you suggested. Consider a signal-detection experiment in which the subject responds with an estimate of the probability that a signal is there. Peterson did this with a single subject over a period of months. If the subject uses probability estimates correctly, then for all responses that the probability is 0.6 that there is a signal, whatever the stimulus situation, a signal should have been there 60 percent of the time if his reporting is correct. So, if you plot subjective probabilities against the relative frequency with which there was, in fact, a stimulus, giving the identity line for reference, you find that the subject's estimates are nearer 50–50 than they ought to be. Therefore, you get an S-shaped function, crossing the identity line at 50 percent. This turns out to be very little different from one subject to another and a highly regular function.

MARKOWITZ: After the observation, you ask him what the probability was that there was a signal.

EDWARDS: That is his response.

MARKOWITZ: I thought Senders was referring to an a priori judgment. Before the observation you ask the subject how well he thinks he will be able to do.

SENDERS: That is right, we might be talking about the Cooper Scale or pilot rating. As he flies the machine, he says: "It is absolutely all right," or "it stinks," or "it is barely controllable." In a sense you are saying that if you asked the pilot to put it in quantitative terms, he might say, "I would guess about 50–50 it could be used for a mission." Would he not say something like that?

BELSLEY: Yes.

MARKOWITZ: So he is making judgments before the fact!

BELSLEY: And after the fact.

SENDERS: To build an airplane in order to find out whether it is a good airplane is expensive and time consuming. We would like to be able to generate analytically a good estimate of what the pilot's prediction will be if you build the airplane. I think that with regard to handling properties this

is what McRuer is doing, and with regard, say, to visual scanning, is what I am trying to do. With regard to total central work, it is what Kristofferson is trying to do. We all want to avoid the necessity of constructing something and trying it out (which is, admittedly, a very good way of finding out whether you can do it, but is also expensive and time consuming).

JOHN A. SWETS: Do you know of any data on the validity of pilots' ratings that used some psychological test? Part of the question is that we want to test in a variety of different ways and we never want to ask the pilots. Are there any well-known instances in which the pilots just do not do very well, where the ratings are not reliable?

SENDERS: The only one I know of was, I think, in 1953 at Wright Field. A nonlinear yaw damper was tried out at the flight test unit for comparison studies of the nonlinear damper versus the linear damper. The pilots took the aircraft up for simulated air-to-air gunnery. The opinions were in favor of the linear damper. The gun-camera records showed that the nonlinear damper was superior. When the records were shown to the pilots, they changed their opinions. In a sense they said, "We were looking at the wrong criteria. We were looking at the ease of handling as opposed to stiffness of point of aim on the target," or something of this sort. In dealing with pilot opinion, there is always a risk that the pilot can be judging on the wrong criteria.

GEORGE A. RATHERT: In those particular tests, you had the problem I spoke about earlier. That airplane was not going to be used with fixed iron sights. It was going to be used with a disturbed-reticle gunsight. The pilots could have subjectively made the judgment of what was best for the airplane with the weapons system they were going to use it with. The pilots will give you the right answers—if you ask the right questions.

If you had asked them which is the best airplane with an iron gunsight, you might have gotten a different answer than if you had asked which is the best airplane, be-

cause they will answer the latter question in view of the use they know you are going to make of it.

SENDERS: I think that the question was essentially to compare the aircraft damper systems. That may be an improper question.

EARL A. ALLUISI: We have gotten off onto the question of asking the pilot about the aircraft, whereas I thought we had started with the question of asking the pilot about himself. Pilots have been asked how well they did or how well they were going to do in some of our work-rest schedules. The question was asked on the a posteriori side; that is, the pilot was asked how well he did. We collected data in a 30-day study with 10 rated pilots from the Air Force. Each subject was required to make an entry in a log at the end of every 2-hour period of work. One of the entries required of each subject throughout the 30 days was an estimation of his efficiency during the preceding 2 hours, where 100 percent efficiency meant (or should have meant) that he had operated at the top level of which he was then capable. The analysis of the data indicated only one significant correlation between this index of efficiency and any of the performance measures. That was a correlation of about 0.30 with performance on the target identification task. I would interpret this result as an indication that the pilots based their estimates more on their performances of target identifications than on the other tasks. In short, in our experimental situation, we can predict performance better using just about any method other than asking the man.

SENDERS: Could you repeat exactly what you asked them to do?

ALLUISI: They were to enter in their book a numeral indicating the percentage efficiency of their work during the preceding 2-hour period. For each man, the baseline (100 percent) would be his top, the highest he was able to perform.

SENDERS: At any time?

ALLUISI: At any time. We had no difficulty getting them to write down numbers, but. . . .

SENDERS: Presumably he was always operating at his top wherever he was, was he not? You were actually asking him to estimate his capacity vis-a-vis his capacity at some other time?

ALLUISI: Did he during the last 2-hour period work at the top level that he could work, or did he work at some point below that; if so, at what point below that? If, during the last 2-hour period, I operated or reported that I operated at 80 percent efficiency, the figure was to have been judged against what I could have done today at that period if I were fresh and in best condition.

JEROME I. ELKIND: May I change the subject a little bit? I want to come back to some of the things you talked about in your test matter. It seems to me that for your results to be really useful to a systems designer, it would be much better if you were measuring different kinds of things from the things you are measuring. Let me give an example. For some of your tasks there are now—although there were not several years ago—fairly good mathematical models of human behavior. For example, if you want to draw a simple model of a human as a continuous tracker, two things are important: time delay and gain. These two things describe how he will behave as a tracker. As a systems designer, I would like to see measures of his gain and his time delay as a function of stress and other factors, rather than measures in mean squared error, because I can predict from the time delay and gain how he will behave in a wide variety of tasks.

The same thing applies to some of the detection tasks, where you might be interested more in d' and the criterion than in some of the more straightforward measures. It seems to me that it is possible to sharpen up the information you are getting out of that type of test and make it much more directly useful to a wide variety of problems.

ALLUISI: To use d', I must have a fairly good proportion of misses, which I do not obtain. We have nearly 100 percent hit rates.

ELKIND: You might have to do a somewhat different experiment.

ALLUISI: I am a little conservative about changing the tasks to match some of the things we have done in other psychological experiments, because I do not want the subject in a game-playing attitude rather than a work attitude. User acceptance is one criterion that has been most difficult to meet in most psychological testing. We have managed to get user acceptance with our battery of tasks. In fact, we have managed to get pilots to work it for 30 days, and they came out essentially saying, "We worked it for 30 days and did not play at it like a game." In order to maintain this acceptance, we must provide the subjects with a realistic task, and to be realistic we do not present signals so brief that most are missed. Realistically, we human workers do not fail and continue working; we cannot keep a subject working if he is continually failing.

The worst subject I have had was a GI, about 70–75 percent accuracy on the arithmetic tasks; the best was an Air Force Academy cadet, 95–96 percent accuracy. I hold that the critical point is in the 80-percent range. If I can hold the subjects at about 80 percent, then I can maintain the work atmosphere and the motivation.

Again, this is one of the grosser variables. We do not do things for 20 or 30 days at which we are continually failing. What we will do is bring up a defense mechanism; if nothing else, we treat it as a game, as if it is not to be taken seriously.

SWETS: You have not spoken of the possibility or the desirability of measuring gain and delay, have you?

ALLUISI: No, I have not, because I do not have tracking in my battery right now. The two that I showed were the initial battery (which was thrown out because of unreliability) and the other one is in the Douglas battery. I would say, yes, I hope to measure gain and delay when we add the psychomotor task, if we add it in with the appropriate measures.

BELSLEY: Suppose you wanted to find out how hard a pilot was working; how close to his capacity he was working when he was flying in an airplane; and how much reserve he has in a given situation? The only mechanism that I know of—or that I am willing to trust at this point—is to ask him. We are looking for any other quantitative measures that can determine what the workload is.

SENDERS: The use of auxiliary tasks has been tried.

BELSLEY: All that does is load the subject to the point where he really stops doing his job. Or, in the landing situation, when the chips are down, he quits looking at auxiliary instruments and only looks at two or three.

SENDERS: In other words, even the entire cockpit is then too much for him?

BELSLEY: Yes. You know what he does, he watches attitude, airspeed, and the altimeter. When he cannot sample those fast enough he must get out. What I am looking for is a quantitative measure of determining his workload on line, other than asking him the question. Alluisi is talking about the same kinds of things, but we cannot necessarily get these data out of real-life situations. This is the fundamental, basic problem.

YNTEMA: Let me object a little to what you are saying. You want to find out what the workload is. One does not want to find out what the workload is just to find out what the workload is. If the subject is performing the task perfectly but is doing so by drawing on his capacity almost to the limit, that is one thing. If, however, he is performing perfectly but has lots of reserve capacity, that is another. They are different because you want to know how much capacity he would have left to deal with another problem if it arose. The direct way to find that out is to give him an auxiliary task to deal with.

BELSLEY: But you do not want to give a subject a nonsense task just to load him up. You want to design a task that will go along with what he is doing so that he is operating at a certain percentage of his capacity.

YNTEMA: Maybe it would be better if your auxiliary task were one that would actually arise in an emergency situation.

BELSLEY: I do not necessarily object to doing this in an experimental fashion, but we cannot do it by applying the results of basic psychological theory; and this is what we are really getting at.

SENDERS: Suppose that a pilot landing an aircraft was instructed to maintain all the relevant parameters, say, with regard to the programing of altitude, airspeed and course along the localizer, and flight path as accurately as he could. If you were to record eye movements and the time functions of error, then I think that logically there would be an expansion of the errors around the zero error line as he approached his limit.

BELSLEY: I agree, but I would tell him: "I want you to be sure to be able to land this airplane. I do not care what you do up until the time you land it." That is the only thing you ask him to do.

SENDERS: You could infer from that what the time course of loading on him was throughout the entire maneuver.

ELKIND: I do not understand. At one time you say you are interested in workload, at another time you say you are interested in whether or not he lands the airplane.

BELSLEY: No. I say that really you are asking the pilot to land the airplane. You are interested in his doing this task. Yet you may put him in the position where he cannot do it. Can you predict this before he lands?

HAROLD G. MILLER: It would certainly be interesting to find out if he can do what you are asking for, because we run it just like you do. If you give the pilot too many calls, he will blow the whole mission.

RATHERT: This is the basic and classical problem in all of our simulator flight test work. The first piece of data we have is a go, no-go pilot statement of the feasibility. We can either ask his opinion or he can show us. But then you ask: "Okay, he did it successfully; but was he right on the ragged edge, or was he doing it well within his capability?" This is the thing we cannot get at. We cannot get a quantitative approach of anything we measured. Sure, you can get a dropoff in the amount of gain the subject can

develop, but he has already gone past physiological limits.

ELKIND: He might still be landing the airplane.

RATHERT: Sure. Pilots have landed airplanes under this condition.

ALLUISI: I think the question that Rathert asked is the real one—the touchstone of whether we are getting there or not. We have two or three problems, and one of them is in the aircraft. We are concerned with safety and, consequently, back off from any real ability to measure. In the simulator, we involve the human and his responses; however, the fact is that he might not let us do what we want to do in order to measure his loading or his capability to handle the system. There are two ways that possibly could be used to approach the performance limits. One is by burdening the subject or pilot with known auxiliary tasks. We might be able to get him to do this by selecting the tasks appropriately; in other words, by using things that he would not regard as a game. Perhaps by giving him a task that is involved in the system itself, we might be able to get real work out of him. The other way is to remove from him some of his normal capability, by either sleep loss or some other stress, then seeing at what point he breaks. In other words, we must either load him up further or remove some of his capability; because, as the system now exists, if he is performing in it and meeting its demands, it is obviously within his capability. We want to find out how much it is costing him to meet the demands of the operations. We can do that only by taking some of that cost and using it for something else; for example, for handling the stresses of sleep loss or of other tasks.

In the early 1950s, when we were running the air traffic control studies at Ohio State, we found that when we put identification on the scope, the controllers could handle more and more aircraft. They started off with the typical procedure of handling three aircraft. If there were more than three on the scope, the controller would order them into a fixed orbit. They would then handle the first three,

put them in a line, and bring them into the ground-controlled approach gate. Once we put an identifictaion of each aircraft on the blip that appeared on the scope, a controller could handle eight or nine aircraft at the same time and bring them in—not by taking them 50 miles out to line them up, but by bringing them directly in and putting them in line about 10 miles from the gate. Among the parameters we wanted to determine for use in our future studies was how many a controller could really handle. So, one night, at the end of the evening's experimentation, we talked one of the controllers into letting us load him up—keep putting aircraft on the scope to see how many he could carry. It had taken about 3 months just to get him to agree to let us try! This was purely simulation; he knew the "pilots" were sitting in the next room operating electronic signal generators, not real aircraft. We had about 17 blips on the scope when the controller called over his radio intercom, "All right, orbit in position." He reached down, cut off the scope, and said, "That is it!" We said, "What do you mean, that is it? You are still handling them." He said, "That is as far as I go." He would not let a midair collision occur on the scope—not even a simulated one! We always have that sort of problem when trying to measure the limits of capability. It would not have been right for us to have put the controller through the emotional experience that a (simulated) midair collision would have given him. It was too much for him.

BELSLEY: That is defined as a realistic simulation.

ALLUISI: In one flight we had a midair collision. The two controllers were really broken up. We had to cancel the experimentation for the rest of the evening. They could not take it.

RATHERT: That is my point. Generally speaking, there are exceptions, but when I can get a pilot to the point where this controller was, I am in medical trouble. I have trouble with my medical supervisor before I ever get the pilot to that point. Perhaps I could do something about this by adding auxiliary tasks and making the whole situation so complex that he breaks down mentally. I am not sure I am entitled to break him down mentally and not break him down physically.

ALLUISI: He will quit before that; he will stop doing those other things, that is what will happen.

BELSLEY: We really need to immerse him in as realistic a simulation as possible or else divide the situation into bits and pieces based on the fundamental psychological data available and predict where the break will occur. This is why we support research in this area: to be able to make this prediction based on the fundamental data.

ALLUISI: I agree fully. That is what we need and that is what we should be working toward. We do not have it today.

RATHERT: May I quickly introduce a change of subject? I like Yntema's taxonomy, but it brings out the solution to the problem of the upset in turbulent air. What is happening in upset and turbulent air? The commercial jet transport pilot was weighing and balancing; the regular NASA or NACA-Ames test pilot had no trouble with the problem. He was not weighing and balancing. In a problem like flying a jet transport in turbulent air, you must determine how to train a man to go from one stage to another. Is this a science? Are there people we can call on who can move in and look at the situation and tell the airlines how to do the retraining? Because that did not happen. This problem involved paying passengers. No human-behavior expert set up a training program that took the airplane pilots quickly, safely, and effectively from one type of decisionmaking behavior to another. Is there a profession that does this?

SENDERS: You say that the NASA or NACA pilot did not do whatever the jet transport pilot did?

RATHERT: He did not get into trouble.

SENDERS: Are you saying that there are two identical situations that occurred or absolutely identical situations?

RATHERT: Within the reasonable limits of airplane moments of inertia, and so forth;

yes. I think the key was that the test pilot is accustomed to unusual occurrences on his ship. We have got him up there with a fully instrumented airplane looking for them.

BELSLEY: I want to speak on this point. When we made a simulation using our height-control apparatus, we simulated the rough air problem so that the airline pilots would say, "Boy, you have got it fixed. It works just fine, just the way it is up there." They could go through the same maneuver. Our pilots sat there and ran through the maneuver and never had a bit of trouble. Then it turned out if you told them—I cannot remember which it was—to follow airspeed and ignore the altitude of the airplane, they flew the airplane in an entirely different way. The airline pilots were flying it purely because they remembered experiences of flying sweptwing jet fighters at these high altitudes.

RATHERT: We thought our simulation for flying jet transports in turbulent air was a total washout. Our test pilots did not get into trouble on the thing. We thought we did not have the trouble. We brought in some regular commercial jet transport pilots and off they went. In one case a pilot lost 35,000 feet of altitude!

SENDERS: Which he was fortunate enough to have.

RATHERT: Right, but when we put our test pilots in the starting environment they straightened the airplane out and flew it. We thought we were a total washout until we brought in some airline pilots. You cannot get a NASA test pilot to have that same experience, it simply would not happen. The reason is that they are on a different level of decisionmaking activity. You have convinced me it is still decisionmaking. Before this morning I would have been tempted to call it reflex behavior. What I want to know is, Is there a professional discipline that analyzes these things?

MARKOWITZ: You practically answered your own question by pointing to the test pilots. The more experience you have with this task, the better you can relegate it to a peripheral system called training. But the required level of training is beyond what an airline pilot might normally go through. The NASA test pilots, as you said, experience these peculiarities all the time. There is a certain method of critical incidents. The ones who did not learn by experience probably are not NASA test pilots any more. So you have a select group with a highly overlearned skill.

On the question of whether you can take commercial pilots and teach them this skill, the answer is probably no—for the same reason you cannot get commercial pilots to do a lot of things that you would like them to do. I bet we could get naïve subjects to do this.

RATHERT: This has been done the hard way, by interoffice memoranda and precept and by sending statistical samples of airline pilots to Johnsville to fly a simulation on the Johnsville centrifuge. The pilots have spread the word. There must have been a simpler and more direct way to accomplish this particular thing.

MARKOWITZ: My last comment is my personal feeling. We could get a nonpilot to learn to fly your simulator in the most severe turbulent conditions that your NASA pilots can and we would have tremendous difficulty in getting a commercial airline pilot to do it.

RATHERT: That may well be.

YNTEMA: In answer to your question whether there is a body of knowledge or a profession that can teach people pattern recognition, I think that there are two places we could look: One, in academic psychology in the field of concept formation or concept recognition of complex stimuli. (One person who has worked on that, using computer techniques to give the feedback, is Swets. Those were auditory stimuli.) Another place is the training work that the Navy has done, in their Special Devices Center. A good deal of what they were trying to do, in retrospect, was to move decisionmaking from category 3, weighing and balancing, into category 2, which is pattern recognition. A subject says, "This is a situation that I have encountered several times before in

the simulator, so I do this action without weighing and balancing."

RATHERT: I did not mean to take this much time, but I had a very practical problem in that the flying safety director of the airline came to us and said, "Okay. You have done a beautiful job with two of our people. What do we do about the rest of them?" We did not know what to tell him.

STANLEY DEUTSCH: Is there not a corollary to this, too, in the area of pilot-induced oscillations? As I recall in this particular case, did you not find that when a pilot went from category 3 to control of the aircraft, that seemed to straighten out the problem itself?

RATHERT: This is true. We had roughly the same experience, but, instead of an average commercial transport pilot, the test pilot was the average young, eager, Air Force fighter squadron pilot. The Air Force itself noticed that they had that distinction.

ELKIND: We did an experiment with naive subjects in which we would change the dynamics of the vehicle quite drastically. Because of the nature of the experiment, the subjects got a tremendous amount of exposure to this kind of a control situation. Several of the subjects got to the point where the whole thing moved up to a sort of satellite computer operation. They could not even tell you there had been a change of dynamics. The whole thing was a preprogramed operation. Furthermore, they, in fact, as your test pilots must have done, developed the skill to handle these kinds of dynamic changes in general. This is a type of a classic problem that they learned how to solve.

BELSLEY: It may well be that the average airline pilot, if you want to call him that, never gets into this situation in flying his machine until it is crucial, and that our people are in and out of this more often. I think this is true; because, at the time, we were having our people fly aircraft that were slightly divergent, so they would know how to control the situation. They had a lot of practice. This makes a difference. In fact, for flying VTOL airplanes, our

people maintain that the best training is to learn how to fly a helicopter. Once you qualify on the helicopter, you come back and fly the VTOL airplane, and then you can do a pretty good job. If they put a pilot right on the VTOL airplane, he has not built up this reaction or knowledge of what can happen to the proper extent. This is exactly the same thing we are talking about.

EDWARDS: I would like to say a little about Yntema's classifications, which I like very much. There was a time a long while ago when Tolman was presenting a cognitive approach to learning and was being criticized, somewhat unfairly but rather imaginatively, by some of his more behavioristically oriented contemporaries. They described his psychology as leaving the rat lost in thought, unable to figure out what to do on the basis of his well-defined cognitive map of what the world was like. That was in some sense a fair criticism of Tolman. I think Yntema's comments underline it. There is a similar sense in which many of these contemporary introductions of decision-theoretic ideas into psychology are leaving the subject similarly lost in inaction. Relatively low-level decision-making processes help provide the human being with a capability. This is especially clear in the case of motor skills, but it also applies to signal detection, learning, and so forth. But these lower level applications of decision theory say little about how a subject will use the capability or for what purposes.

There are at least two different ways in which linkage between lower level and higher level decision processes occurs. One of them is that a person typically uses a skill—perceptual, motor, or anything of the sort—in the service of some goal. That goal specifies signal-detectability theory where you are going to put your criterion. In some tracking models, it specifies how tightly you wish to control. And this is built into the theories. That is, you find these goals in them, but they tend to float in midair, and it is not clear how they link with the larger goals being served by the skill.

The other linkage, which is in some ways more difficult and which I do not think has been discussed at all, is that perception of one's own sensory and motor capabilities has a lot to do with the weighing and balancing kind of decision. Consider Tanner's example of the landing approach. Knowing how incompetent I am at following that localizer, I take my sensory-motor incompetence very much into account in deciding how low an approach to make. I apply a personal correction to the FAA minima. I do not think I am alone in this either. I am attempting to say that there are linkages between the third type of decisionmaking (which is, of course, what is traditionally meant by decisionmaking) and these other varieties. Psychologists have produced very, very little discussion and very, very little theorizing about the nature of these linkages.

ELKIND: I agree with you. As a matter of fact one of the things I wanted to talk about tomorrow was where this type of linkage was direct and explicit. I think, for example, that the problem that George Rathert talked a lot about, recovery from unusual flight situations, can at least crudely be put into a fairly standard decision-theory framework. I think that that helps to crystallize some of the training problems. It also turns out, if you look at modern control theory as it is now being practiced, that there is, unlike the sort of conventional control theory, a very explicit statement of what the goals of the control are. There is a well-defined performance functional, and this is to be maximized or minimized. And the relationship between the control behavior and the performance are well established.

EDWARDS: What about the relationships between the control functional, which I suppose means a utility function of some kind, and the broader setting in which the work is being done?

ELKIND: I am not quite sure how you define that broader setting, really.

EDWARDS: What is the man trying to do? What are his purposes?

ELKIND: In a narrow sense, I think this is one of the purposes in the assessment of current or modern control theory. That is to say, you can talk about problems in which you want to minimize your root mean square error at touchdown. You do not care what you do in between, but the terminal value must be well controlled. Now, you can solve theoretically for the optimum control that will achieve that; you can compare that with human behavior and get something that is appropriate. That is the type of thing that I think you have in mind.

YNTEMA: Part of what you are saying is that when you get into a weighing and balancing situation or into a choice of SOP, which is really a short circuit for weighing and balancing, that you must consider payoffs. I am wholly with you. The particular example I was talking about did not involve weighing both probabilities and payoffs, but in many of the real-life decisions, that is the essence of the problem.

EDWARDS: I was saying something different from that, although I certainly would agree. I was inquiring about the linkages both up and down between the different levels of intellectual function that you were talking about and was suggesting that these linkages are themselves both important and interesting topics of study about which remarkably little has been done. Elkind was talking about one class of such linkages. Senders mentioned another example in which a driver in a simulator judges whether he can get through an obstacle, and thereafter drives through it. In that case, of course, you compel him to drive through it. In a more real-world task, he might have an option, such as to go by some other route. At that point there starts to be a clear-cut linkage between his judgment of essentially what his characteristics are at level 1 and his behavior at level 3. I think there are lots of these kinds of linkages, and it seems to me worthwhile to start asking questions about them.

ELKIND: I really do not see the difference between the three levels, at least not in the way I think you see it. It seems to me that there is a temporal element

dominant there, that certain things happen very fast and weighing and balancing will take a long time.

BELSLEY: That is known as weighing, not balancing.

ELKIND: Deciding whether you are going through an obstacle as you are approaching it, whether you will drive through or put on the brakes, looks to me like a fairly automatic thing. I doubt that you have time to do much weighing and balancing.

EDWARDS: As I understand what Yntema said about the satellite computer—we are in trouble on that score too. The defining characteristic of level 1 is that you have no access to information about the process.

YNTEMA: That was the point. You have no conscious access to these decisions.

EDWARDS: I claim that in such situations I can do a retrospective-introspective analysis of what it felt like, what the considerations were. That is why I disagree with this long-time idea. It is easier when it is a long time, but I claim you can do it anyhow.

ELKIND: I am not sure I can concur.

YNTEMA: One could say these things a litle more conservatively. Tasks in the No. 1 category can go on without any linkage to consciousness. You can be carrying on a conversation while you are doing one of these things. Maybe it is too strong to say that they are totally unavailable to consciousness, but introspectively there are some very skilled activities, it seems to me, which involve small-scale decisions that are completely unavailable to consciousness. You cannot get them back any more.

EDWARDS: I would propose that the defining characteristic for level 2 is that you recognize the right answer, that there is a right answer, and that you can be sure without waiting for the results of the actual behavior that you have it.

ELKIND: That may also be something that happens that is not unconscious.

EDWARDS: No. He was talking about levels 1 and 3. I am talking about the distinction between 2 and 3. The distinction between 2 and 3, it seems to me, is that 2 is essentially what they call decisionmaking under certainty; that is, there is no uncertainty as to what the situation is.

Simulation Training for Flight Control Decisionmaking

HAROLD G. MILLER

Manned Spacecraft Center

Throughout Projects Mercury, Gemini, and Apollo, flight-control personnel have manned consoles specifically for monitoring and controlling the activity of the flightcrew and the spacecraft systems. The flight-controller function has progressively increased in degree of difficulty as the spacecraft and the ground system have become more complex. The simulation training he receives in a mission environment has been the final hone that sharpened the flight controller to the state of readiness necessary for a successful mission, even under the most adverse conditions. Training and simulation objectives are directed at developing proficiency and competency in the flight controllers to perform the mission-support function. To understand the importance of simulation exercises to mission success, we must understand the duties of a flight controller, both prior to and during manned space flights. We also must understand how the flight controllers are organized and with whom they interface in the decisionmaking processes.

Flight-controller training encompasses the entire spectrum of the space program. Training begins as soon as he starts to work, and consists of formal, intrinsic, and simulation training. The training culminates in a final series of simulation exercises for the flight controllers to insure that they can handle specified contingencies and verify the procedures they have established. Simulations also establish confidence in the flight controller of the system he must use during the mission. Of significant importance is the value the simulation system has in developing mission rules and in determining the readiness of the mission facilities. This paper will describe the functions and duties of a flight controller and how the simulation allows him to make decisions.

The flight controller is a planner, implementer, trajectory and systems expert, operator, and a decisionmaker. The flight controller is a planner heavily involved in the area of planning of the missions and his console procedures. In his mission planning, he must specify mission rules and timelines that will govern his actions throughout the mission. He must plan aborts to insure that, if a contingency occurs, procedures are available to insure crew safety. The flight controller, in conjunction with the crew, must determine the flight plan that is to be followed, and it must account for items such as experiments, sleep cycles, location of manned space flight network (MSFN) sites for critical activities, such as orbital maneuvers, and so forth. Also, the flight controller must determine what data are to be transmitted from the remote site to the Mission Control Center (MCC) in order that a proper evaluation of the spacecraft systems can be made. The planning continues in an iterative cycle until the mission occurs, at which time the flight controller must implement the plans he has made. The plans involve the use of mission rules, detailed test objectives, abort plans, flight plans, and so forth.

The system and trajectory of the spacecraft must be known and understood by the flight controller. He must be intimately acquainted with their limits and their constraints so that as he performs his flight-controller duties in the mission control room, he can make decisions concerning crew safety. This understanding is mandatory because the many variables in the spacecraft and the ground systems preclude the possibility of documenting all but the most basic alternate procedures. He must listen to information from different sources and react to these inputs, sometimes instantaneously.

Pressure is ever present with a flight controller while he is doing his job. For instance, during the AS–204L mission, about 97 percent of the mission objectives were achieved in spite of the fact that during the first descent-propulsion-system burn the lunar module (LM) guidance computer functioned improperly, and, in order to overcome this failure and achieve the mission objectives, the flight controllers had to assume the computer's function for a period of 5 hours.

Finally, the flight controller is a decisionmaker. No matter how good a planner or implementer he is, how good a systems man he is, or how good an operator he is—he must develop the ability to make decisions in real time at the console in order to qualify as a flight controller for a mission. What kinds of decisions does he make and what actions does he take? During a mission, the flight controller is diagnosing the spacecraft and trajectory continually. If an anomaly is suspected, the first action is to determine the validity of his data. Once he decides that the data are valid, he must take actions to correct them, and he must make a decision as to what course of action to take. The ease with which he makes these decisions is a measure of the thoroughness of his planning and the completeness of his training.

Figure 1 represents the organization of a typical operations team supporting a lunar manned space-flight mission, showing all of the flight controller positions. The Flight Operations Director is responsible for operation, mission planning, and the overall direction and management of flight control and recovery activities associated with real-time mission progress. The Flight Director is in direct command of the other flight controllers and is responsible for their real-time decisions. In case of an emergency, he can terminate the mission. The Assistant Flight Director acts as a staff assistant to the Flight Director. The Flight Director's prime support would be the spacecraft communicator group, the flight dynamics group, a booster and spacecraft systems group, flight surgeons group, and the experiments activity group. Each function is supported by specialists located in support staff rooms equipped to permit detailed systems analysis of the spacecraft, the crew, the experiments, and so forth. In addition, there are communication circuits that enable the staff support room personnel to discuss problems with other specialists; for example, the design engineer, who may be located at some remote facility.

During a mission, the flight controller interfaces with the flightcrew, support personnel, other flight controllers, and his console equipment. Prior to the mission, the flightcrew and the flight controllers must work together to establish their operating procedures and to determine what data must be transferred between the flightcrew and the flight controllers for each phase of the mission. The critical aspect of this interface is during the flight when you have one person on the ground (the capsule communicator) talking to the crew. He is usually an astronaut. All communications, except for emergencies, to the crew in the spacecraft are relayed through him. The importance of this interface is that during critical phases of the mission, such as during launch or some of the maneuvers, you must get the right words, at the right time, in the right order, to the astronaut in the spacecraft. He must know what to expect in terms of data that are coming to him. The ground personnel, in turn, must know the kinds of data they are going to receive
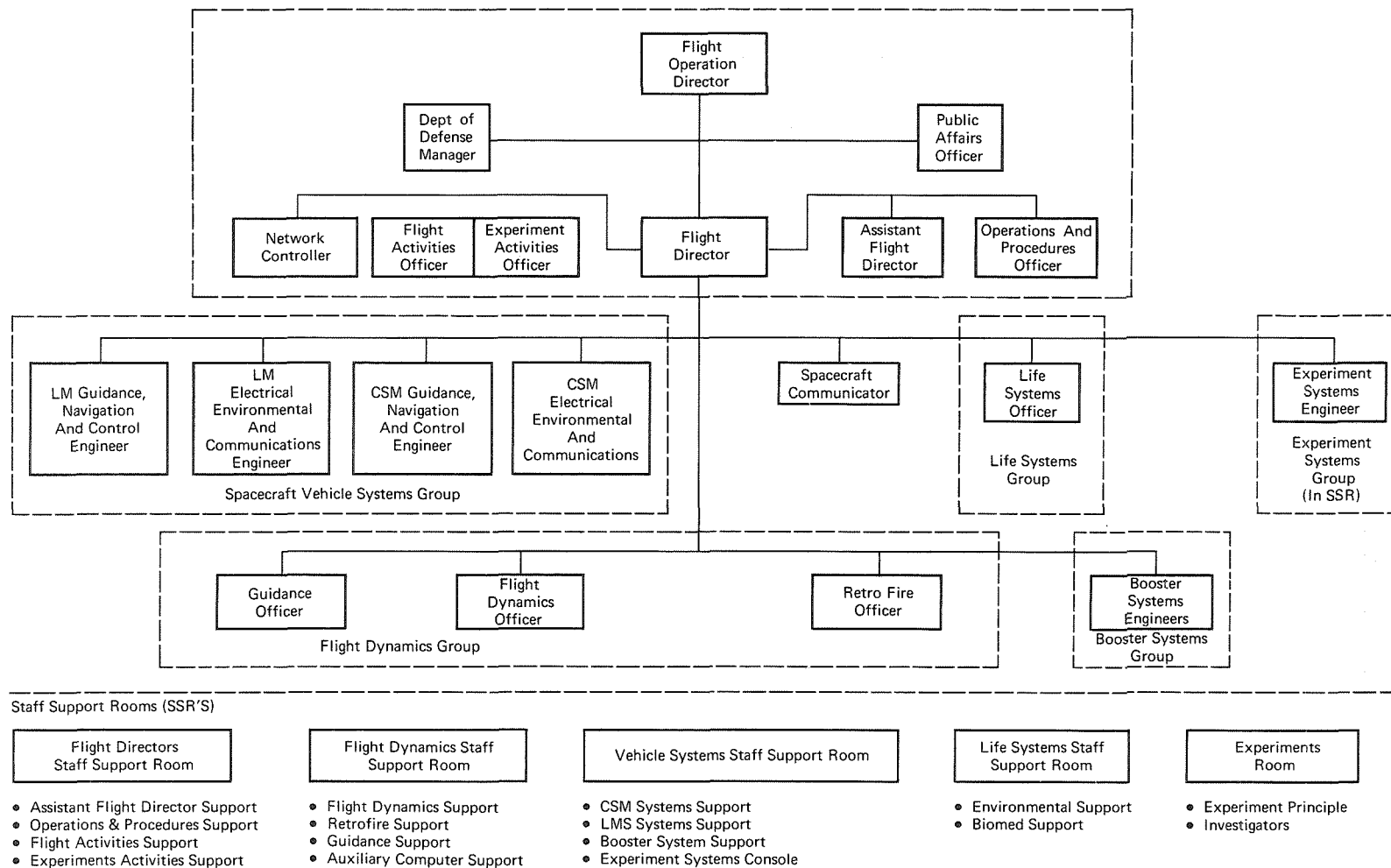
FIGURE 1.—MCC–H mission operations control room organization and staff support room interface.

from the crew. These communications are a matter of timing. The entire procedure is as precise as a fine clock. The simulations conducted prior to the mission stress the exercising of these procedures.

Under certain conditions, the flight controllers must interface directly with the technical support personnel who support the mission but do not have flight control responsibilities. Such conditions exist when changes to spacecraft instrumentation have been made and the control center was not modified to display the information to the flight controllers. The technical support personnel operating the ground equipment will have access to these data that are needed by the flight controllers. These interfaces become extremely important when having the maximum data is mandatory for resolution of a spacecraft problem. The simulation exercises test these interfaces and establish a repertoire between the flight controllers and the technical support personnel so that information can be rapidly obtained by the flight controller.

The development of procedures for the relations among the individual flight control team members is a prime objective of the simulation exercise. Each team member has both his individual action he can take and those in which one or more other flight controllers are involved in accomplishing a task. The flight controllers develop documents such as the Flight Control Operations Handbook, individual console operating procedures that detail these actions down to the lowest level. Some of the more complex interfaces are associated with the generation of commands to be transmitted to the spacecraft.

Through his console, the flight controller has access to communication loops, TV displays, control switches, event status lights, and other devices that allow him to evaluate the spacecraft and the ground network support systems. The flight controller must be completely familiar with the capability afforded him by his console. This can only be accomplished through extensive use of the console in as close to a mission environ-

ment as possible. This flight-controller training environment is provided by a variety of mission simulations.

The overall scope of the training a flight controller receives and must successfully complete prior to qualifying to support a mission is shown in figure 2. Three types of training are used: formal, intrinsic, and simulation. The formal training consists of lecture courses. Work is being done in the area of programed instruction. At present, we are developing programed instruction courses for spacecraft systems and for the manned space-flight worldwide network. Additional courses are developed as needed. Such courses might cover the operation of research equipment to be used on a particular mission. Practical exercises and field trips to contractor plants are also provided.

Flight Controller Training

Formal	Intrinsic	Simulation
System Courses	Personal Contact	Unit Training
Specific Course	Design Reviews	Mission Simulation
Field Trips	Mission Planning	
Practical Exercises	Operational Planning	Network Simulation
	Procedural Development	
	Mission Experience	
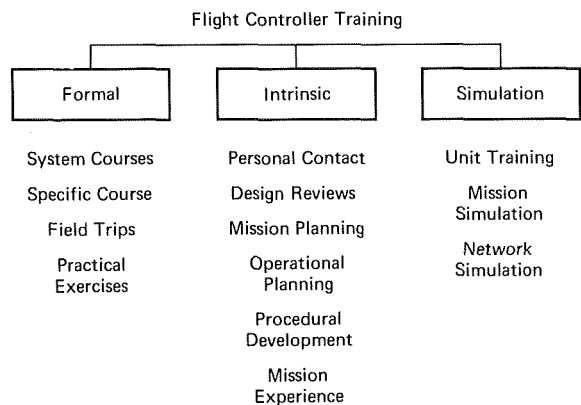
FIGURE 2.—Overall view of the training of a flight controller.

Intrinsic training is the on-the-job training that new flight controllers receive from sitting in the office with experienced flight controllers who work on the same job, and from monitoring missions.

Three types of simulation training are used: unit, mission simulations, and network simulations. Unit training provides exercises for groups of flight controllers, such as the flight dynamics group. This training usually starts about 45 to 60 days prior to the mission and requires about 40 hours per group. In general, these exercises do

not include the flightcrew except where special flightcrew part task training devices are used as the data source for flight controllers. One example of this is the simulation exercises performed with the mission evaluator located at North American Rockwell in Downey, Calif. The mission evaluator is a simulator used by the crew and a group of flight controllers to develop procedures for operation of the onboard spacecraft computer.

The mission simulations, which start after the unit training has been completed, generally include all of the flight-controller groups in a concerted operation and are divided into mission phases. This type of simulation consumes most of the time spent training prior to each flight. In general, 20 to 30 days of total team mission simulation exercises are required for each flight. Approximately two-thirds of these exercises are performed in conjunction with the flightcrew.

Finally, about 5 to 7 days before the mission, the network simulations are conducted. For these simulations, magnetic tapes of recorded spacecraft data are provided at all of the sites all around the world, and are played back in a time order sequence during the simulations. The primary objective of network simulations is to train the flight controllers in as complete a mission environment as possible by exercising the total manned space-flight network as planned to be used for that flight.

Training ground-control personnel involves much more than systems training. It means training flight controllers to operate as a team. They may be the best systems engineers in the world, but if they have not developed well-coordinated procedures for working with each other, they are not ready to support a mission. Of utmost importance is the work on the interfacing of flight controllers with the flightcrew. This is done by interfacing the flightcrew's training equipment with the control center for a complete closed-loop simulation. In this manner, with the crew in their mission simulators and the flight controllers at their con-

soles, the simulation is operated identically with a real mission.

Another objective of simulations is to exercise the operating procedures of the flight controllers and the total MCC. It is important to exercise the interface of the MCC maintenance and operations personnel with the flight controllers, so that equipment problems can be resolved quickly.

Through the simulations, the controllers develop confidence in the systems and computer programs that will be used during the flight. This in no way substitutes for the rigid verification testing of the computer program. The mission operations programs for a typical mission exceed 500,000 instructions and are contained in one computer in the real-time computer complex (RTCC). Throughout the Mercury, Gemini, and Apollo programs, it has been found that, in spite of extensive testing performed to find faults and weak points in the program, when the flight controller operates his console and the system is loaded down, it usually contains some anomalies the first time. This fact does not reflect on the quality of the program testing, which is very detailed and very complete; the problem stems from the fact that flight controllers will do things that the system was not originally designed to permit. The objective, of course, is to have the flight controller feel that the system he has will support him in the mission environment. Our approach to simulation is to make maximum use of the operational facilities so that we can validate the systems operation, the computer programing, and so forth, as configured to support that mission. Use of the flightcrew-training facilities as a data source is also necessary to meet the training objectives. As noted earlier, this allows the crew and the flight controllers to work together and establish their procedures and operational problems prior to the actual mission. For additional flexibility in simulation schedules, we have augmented our data source capability with mathematical models of the command and service module, the lunar module, and the Saturn launch vehicle. This computerized simulation of

major flight systems provides a completely independent system for training flight controllers. Mathematical models for the network exist, so that input data into the control center are representative of data as they would come from the MSFN. And finally, by using operational facilities, we keep the equipment for simulation facilities to a minimum.

Figure 3 is a functional diagram of the simulation system and how it interfaces with the MCC. To meet one of our objectives, the simulation system must input data to the MCC in the format and rate identical

with that received from the MSFN. In the same manner, the simulation system must also respond to all data transmitted from the MCC. The block labeled "simulation equipment" (fig. 3) represents the functions performed by the Goddard Space Flight Center and part of the Kennedy Space Center. Data to and from the MCC are switched into this equipment. The simulation computer and flightcrew trainers represent the airborne system of the spacecraft and booster. Simulation operators man consoles and monitor the MCC data and control the activity of the simulation facilities. The simulation facilities required for Projects Gemini and Apollo represent an outlay of approximately $10 to $15 million.

The organization of the simulation controller is shown in figure 4. Simulation controllers are located in two areas—each representing different functional jobs. One group is located near the flight controllers in order to monitor their activities. From this vantage point, the simulation controller can determine if the simulation is proceeding normally. For instance, an exercise may contain several faults that require flight-control activity; if, however, it turns out that the flight controllers are not handling the planned faults properly, or are having trouble with a procedure, the simulation controller can
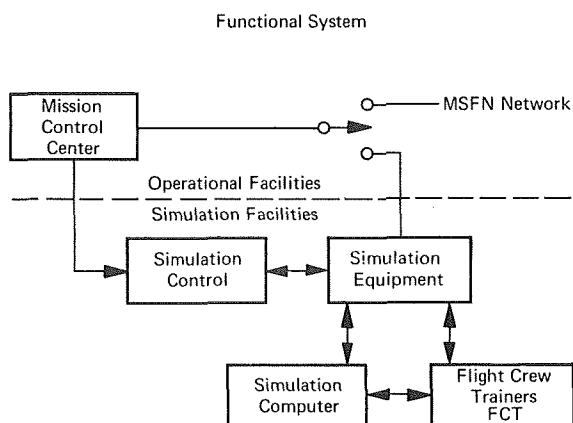
Functional System

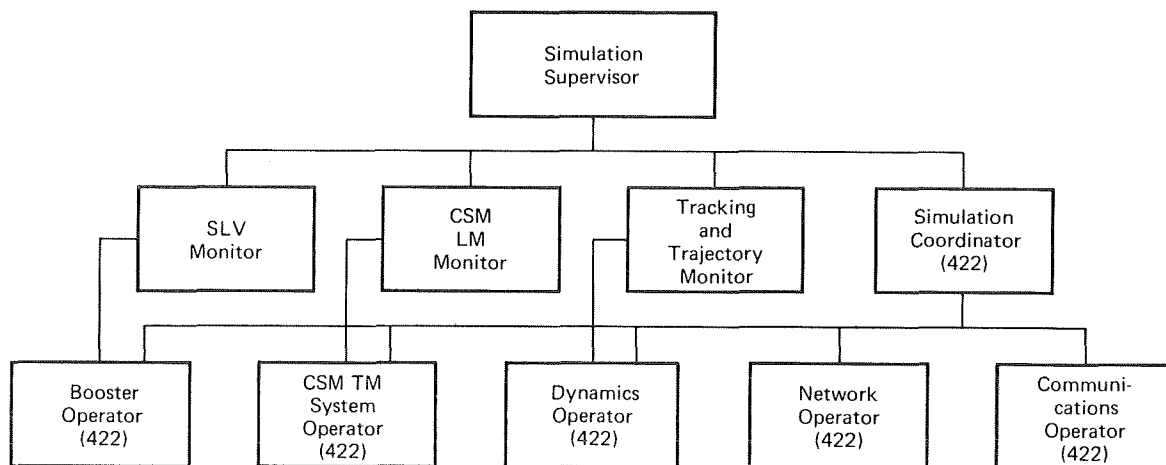FIGURE 3.—The simulation system.

FIGURE 4.—Operational organization for the simulation controller.

alter the flight situation by either eliminating, changing, or adding faults in order to insure that the flight controllers obtain maximum benefit from the exercise. Practice and experience have shown that too many faults in an exercise can result in very little training of the flight controller. This is especially true in the first few exercises conducted for a mission. On the other hand, as the flight controllers progress through the simulation program leading up to a flight, they develop the competence to handle more and more complex problems. Thus, the simulation supervisor must decide when to either add or eliminate faults. This decision is based on inputs he receives from the other simulation personnel who are closely monitoring the flight controller activities. Once a decision has been made to change the faulting scheme for an exercise, the simulation controllers make the various necessary entries into the simulation computers and insure that the changes are effective. The simulation controllers have the responsibility of making all of the necessary computer entries and of insuring that the simulation system is working properly.

Scripting for the simulation exercises requires a minimum of documentation because the simulation system is designed to use mathematical models to represent spacecraft and booster operations. Therefore, a fault entered into the mathematical model results in a correlated set of data for evaluation by the flight controller. This ease of documentation permits much flexibility and speed in developing the most effective exercises for the flight controller. Because the simulations are used by the flight controllers to help validate, and in some instances establish, mission rules, any scripts written prior to the start of simulations may become invalid as a result of changes to the mission rules derived from the simulation. Therefore, the flexibility afforded by the simulation system allows the simulation controller to rapidly adjust the exercises as needed,

either just prior to the simulation or, more importantly, during the actual exercise. A typical example of a portion of a simulation exercise script is shown in figure 5.

Although the documentation for simulation scripts is relatively simple, there is a voluminous amount of documentation that the simulation controller must review prior to their development. A typical list of categories of such documents includes accurate system drawings, operational procedures, mission rules, trajectory papers, flight plan, network documentation, and control center documentation. The documentation that the simulation controller must publish months before the start of simulations in order to provide the necessary training of the flight controllers includes scripts, simulation configuration documents, simulation operations plans, flight operations plans, and simulation requirements.

The simulation system used today to train flight controllers is in concept the same as that used in Mercury and Gemini. Although a digital mathematical model of the Agena was used in Gemini, extensive use of mathematical models for independent flight-controller training only began with Apollo. Throughout the programs, the flight controllers have become much more sophisticated in their handling of each mission. The simulation systems in response have become more complex in order to meet the needs of the flight controllers. Various changes to mission profiles late in the program have dictated a simulation system that is flexible and has a fast response in order to provide accurate training. The simulation system that is in use was designed to these goals and has adequately met these requirements. It has proved itself as an effective training system, in addition to verifying the hardware and software systems within the control center. As future spacecraft and ground systems increase in complexity, the training system used to exercise the flight controllers must grow with the various programs.

Simulation Exercise Script Continuation		
1. Exercise No. 501/LS/005b	2. Date 4—20—67	3. Page _____ of _____ Pages
4. Fault Summary		

SEO/S GET Fault I D Sy/Se Loc	Description

(a) 0:00:15
GN—005
Sy 8
ADRK
	Drift in Pitch Axis of .15 degree per second in IMU.
	SYSFLTCC, GN — 005, . 15, , , , , ;
	Cause: Excessive launch phase dynamic loads cause Pitch IRIG
	              to output erroneous signals to the Stab loop.
	Result: IMU drifts in Pitch axis at rate of .15 degree per
	              second. FDO will send RTC 12 with G&N "no go". MR4—8B.
	Cues: GO521 MO683 AGCU and IMU Pitch attitudes differ
	Backup: GN—007 AGC Pwr Supply Fail
	              Sy 9 ADRK ASAP

(b) 00:03:00
PPS — 416
Sy 1
ADRK
	S—IVB Main Low Valve Fails Closed
	
	Before Engine Ignition
	
	Cause: The mainstage solenoid fails in a deenergized state.
	Result: S—IVB will fail to ignite when the Early Staging
	              Command is sent. FDO will command Abort. MR4—5A.
	Cues: GO305 and MI405. Parameters G3—401 and DI—401.
	Backup: PPS—412, Lox Prevalve Fails Closed
	              Sy 3, ADRK ASAP.
	              PPS—413 Fuel Prevalve Fails Closed
	              Sy 4, ADRD ASAP.

(c) 00:05:00
PPS — 206
Sy 2
ADRK
	S—11 Fuel Ullage Pressure Leak
	SYSFLTCC, PPS — 206, , 20.0, , , , ;
	Cause: Fuel ullage gas (gaseous hydrogen) leak
	Result: When fuel ullage pressure drops to point where fuel
	              pump inlet pressure is less than 20.5 psia, all 5 engines
	              will go out (must be between 00:03:06 and 00:07:53).
	              BSE #1 will send Early Staging Command. MR5—12.
	Cues: GO304 and MI402. Parameters D252—219, D253—219,
	              D13—201, D13—202, D13—203, D13—204, D13—205

FIGURE 5.—Portion of a simulation exercise script.

DISCUSSION

STEVEN E. BELSLEY: I wondered if you might say something about how this system design evolved.

HAROLD G. MILLER: You mean for simulation?

BELSLEY: Well, your overall mission control system evolvement. Your simulation is a representation of it.

MILLER: You had the same people that worked on Mercury working on Gemini and setting down the requirements for the ground system. The Gemini system and the Apollo system, simply stated, are overgrown Mercury systems. We replaced meters with cathode ray tubes. Instead of having cross-amplitude telemetry ground stations, we have pulse code modulation telemetry ground stations. The system got bigger.

BELSLEY: But during the first part of Mercury

there was a tremendous development of their concept, was there not—from the suborbital flight on up?

MILLER: The system changed very little. When you talk about system, are you talking about the actual hardware or the system used for control?

BELSLEY: I am not talking about hardware. I am talking about the overall systems concept of control. I wanted to find out whether this system was designed by the operational people or whether it was designed by some psychologists to work to a set of specifications.

MILLER: If you had asked me that, I would have told you I do not know.

GEORGE A. RATHERT: The technology used would be of interest in other grossly similar problems; air traffic control for the supersonic transport (SST) immediately comes to mind. This technology does not seem to be documented. Admittedly when you start charging through these missions, the mission time schedules are of overwhelming importance and the documentation never catches up. But there must have been a large body of research or at least a large amount of thinking done which would be very interesting.

MILLER: The Mercury system was very simple. You had 88 parameters coming from the spacecraft. You had capsule systems monitors, aeromedics, a flight dynamics officer, a network controller, and several other guys. You had 15 to 20 meters on each console.

RATHERT: But look at how many parameters are coming down from the SST. As I say, the technology is not being applied. So there may be 30 parameters instead of 88, but the technology is not being applied in the same way.

MILLER: I do not know. I do remember one proposed design for the Mercury control center that was simply two men and a red light. Everything would be in a computer. If something went wrong, a red light would flash and the man would push the switch. Proposed configurations certainly did run the gamut, I can tell you that much.

LLOYD A. JEFFRESS: A lot of the communication in these flights is line of sight. Why do we use amplitude modulation for voice when frequency modulation is so much more intelligible? I suppose somebody in

his infinite wisdom made that decision, but I would like to know why, if there is a reason.

MILLER: The Soviets are on 20 megacycles, and they transmit all the way around the world. We use ultra high frequency and line of sight only. I really could not tell you. The quality of communications is much better, I do know that.

JOSEPH MARKOWITZ: I would wager that if you had started out with the simple two-man, red-light, big-computer system and began to exercise people on that system, then these people, being professional and having a certain professional pride, would gradually request more and more explicit information.

MILLER: The system grew. Designed it was, grow it did. How else would you say it? As the people became more familiar with the control system, they started saying, "I need this capability or that capability." Between Mercury and Gemini the system grew some more. There were more data coming from the spacecraft and more flight controllers. Apollo just compounds the situation again, but we are doing it the same way.

BELSLEY: There is a pervading influence that you have not talked about: the ability to identify problems and effect solutions while the spacecraft is in orbit. If you were to take the go, no-go position, that all you had to do was abort, you would never have gotten any place. At all times in this whole system, you can identify the fault and try to rectify it without scrubbing the mission, is that not true? This is why you have got this complete set of data retrieval systems at your beck and call.

MARKOWITZ: In point of fact, any of these options could have been exercised by an automatic system.

MILLER: In 1959 you would have been pushing it.

BELSLEY: I do not think you can do it now.

MARKOWITZ: That is exactly why the mission control console evolved as it did, because every one of the controllers and personnel voiced exactly that argument. And because there is no real counterargument, then, in fact, these pieces were added. In every case, the request for more explicit information came not from the planning people but from the operations people.

Controller Decisions in Space Flight[1]

WARD EDWARDS

Engineering Psychology Laboratory
University of Michigan

The basic idea of contemporary decision theory is so simple that it is trivial. Every decision depends on the decisionmaker's subjective, personal answers to two questions: What is at stake? and What are the odds? The answer to the former question requires construction of a payoff matrix and measurement of utilities. The answer to the latter question requires either direct estimation of probabilities, or, more often in real-world decision, the kind of information processing for which Bayes' theorem is the optimal mathematical model. Use of these subjective stakes and odds to determine decisions requires application of the subjectively expected utility (SEU) maximization model.

This set of ideas has been developed substantially on the basis of laboratory experiments with college student subjects, nickel-and-dime stakes, and pinball machines or bookbags full of poker chips as random devices around which subjective odds estimates must be structured. The claim to generality is made by almost all of us who work in this field—indeed, you have heard several versions of it already at this meeting. But instances of application of these ideas to real-world decisions are rare, and not especially convincing.

When Huff of the Ames Research Center of NASA told me that I could have the opportunity to study real decisionmaking

in the mission operations control room of the Manned Spacecraft Center (MSC), Houston, I was therefore delighted. When I discovered what MSC is like, my delight increased beyond reasonable bound. For one thing, every middle-aged scientist who hides under his graying hair and sagging waistline the soul of a teenaged science-fiction addict has no choice but to be thrilled at MSC, not only because MSC's business is to get men into and out of space, but also because all the fantastic resources of the latest and most expensive technology are so obviously and lavishly used to do so. I must have seen at least 24 computers there, each one as large as any I have seen in a university computing center. Most are on line, attached to delightfully science-fictional consoles that have pushbuttons and flashing lights and steady lights and meters and two TV screens each; there must be 70 or 80 such on-line consoles at MSC.

After I recovered from the initial shock of recognizing a world of which (I have known ever since my teens) I have to be a part, I found more sophisticated reasons for delight. One of them is here now, listening: Harold Miller, Chief of the Mission Simulation Branch. Miller's job—indeed, in many ways the job of most of MSC—is to train mission controllers. More specifically, he prepares and operates simulated missions. The actual men, actual control room, and actual controls used in real missions are used for these simulations; the rest of the

world, including spacecraft, ground stations, communications, and so forth, is Miller— plus a team of 15 or 20 others, plus several large computers and, of course, lots of consoles.

Mission controllers sitting in Houston have little real work to do during a mission—as long as everything is normal (or nominal, in MSC jargon). If, however, something goes wrong, then the mission control room starts to hum like a disturbed beehive, as the controllers strive to determine what is wrong, how it can be fixed or coped with, and what to do about it. Naturally, therefore, the part of their training with which Miller is primarily concerned is response to emergencies, and each simulation consists of one emergency piled on top of another. So it seemed reasonable to Huff and to me that emergencies designed to meet our experimental needs could, without detriment to the purpose of these training sessions, be inserted into certain simulations. Miller agreed and turned some of his most able lieutenants loose on the task of working up appropriate cases, and his various bosses and colleagues, including, especially, the key man to be studied, Eugene Kranz, the Flight Director, agreed also.

In order for you to understand what we did and what was found, I must go into a fair amount of technical detail about the launch phase of the Apollo AS–204L mission. To make it more difficult, during months of working with this material, I have become brainwashed. At MSC they rarely speak English; they prefer a sort of alphabet soup of acronyms mixed with a great deal of jargon. Like a schoolboy showing off his French, I find myself talking that way, too, when I talk about MSC. To help you out, I have included a list of acronyms and their definitions. It may cover the terms I will use—but I can guarantee you that it does not exhaust MSC's supply. About 18,000 acronyms can be made from all possible combinations of one, two, and three letters, and I doubt that MSC has missed many.

At Miller's recommendation, our cases concentrated on the launch phase, before too many unforeseen controller responses to previous malfunctions could foul up our expectations about what might happen. Moreover, they were for unmanned missions, because it was decisions by controllers, not by astronauts, that we wanted to study.

One of our two cases, the less successful one, concerned a leak in a helium tank, a malfunctioning reaction control system thrustor, and a premature cutoff of the Saturn booster. The background goes like this. In a normal mission, after insertion of the booster and everything attached to it into orbit, the lunar excursion module (LM) is detached from the booster preparatory to the module tests for which the 204L mission is intended. To achieve this separation, locks holding the LM to the booster are released, and then some very small rockets called the reaction control system (RCS) thrustors are fired to move the LM away from the booster. These same RCS thrustors are also used for later attitude control of the vehicle. Normally, four of them fire during this separation. Of these, two are driven by a fuel and oxidizer supply system called the A system; the other two are driven by the B system. A crossfeed valve permitting A system fuel and oxidizer to go to B system thrustors (and vice versa) exists, but is closed except in emergency. Associated with each system is a high-pressure helium tank; the helium forces the fuel and oxidizer out of their tanks and into the thrustors. Thus, zero pressure in, say, the B system helium tank would put all B system thrustors out of action, unless the crossfeed valve were opened.

If the Saturn booster stops firing before orbit is achieved, the time pressure becomes intense. Such a case is called a launch abort and, depending on the particular conditions that prevail, a number of different things can happen. In this one, the abort occurred at a time when orbit clearly could not be reached. In that case, there is a rather brief period of free fall outside the atmosphere, 3 minutes or so, and then the spacecraft burns up, or, at any rate, its signal is lost due to reentry. So the goal

in such a case is to perform as many of the tests for which the mission is intended as is possible during the short time available. There are two ways in which such a suborbital sequence (SOS) can be controlled. One is the LM guidance computer (LGC), the computer carried in the LM. This is the preferred way. If for some reason that cannot be done, an alternative is to use a tape reader also carried in the LM, called the program reader assembly (PRA). The PRA can command functions in proper sequence and timing, but cannot modify its actions contingent on their outcomes. For reasons that I do not understand, the SOS under LGC control can be initiated only if the initiating command is received within 60 seconds of the premature booster cutoff; after that, the LGC will not listen to commands from the ground.

With that much background, let me describe the details of our first case, as shown in figure 1. At about 1 minute 30 seconds after liftoff, the system B helium tank de-

---

SAMPLE SIMULATION EXERCISE SCRIPT		
1. Exercise No. 501/LS/005b	2. Date 4-20-67	3. Page 1 of 3 Pages
4. Setup (Fault init., Timing, Modes, Vectors, etc.)		

Fault Init: a)   1. PPS – 416   6. PPS – 214
                2. PPS – 206*   7. PPS – 235*
                3. PPS – 412   8. GN – 005*
                4. PPS – 413   9. GN – 007
                5. PPS – 202

         b)   Contingency Group 1

 * Change value by MED

 Modes/Timing: Ten minutes prelaunch. Real GMT.

5. Exercise Summary/Objectives

 The objective of this case is to cause a Mode 2 abort condition with attitude control for proper entry attitude. Booster initiation of early staging command is exercised by early cutoff of the S-II stage. The Mode 2 condition is caused by the subsequent failure of the S-IVB ignition. A drift in the IMU pitch axis will force failure of the G & N by command by the GNC.

6. Comments

 VAN acquisition will be early and short, if at all. A possibility exists for a short SPS burn if the G&N is not failed. If this does happen, the SPS will be shut down by FIDO using RTC 12.

FIGURE 1.—Sample simulation exercise script.

SIMULATION EXERCISE SCRIPT CONTINUATION		
1. Exercise No. 501/LS/005b	2. Date 4-20-67	3. Page 2 of 3 Pages
4. Fault Summary		

SEQ/SGET Fault ID Sy/Se Loc	Description
(a) 0:00:15 GN — 005 Sy 8 ADRK	Drift in Pitch Axis of .15 degree per second in IMU. SYSFLTCC, GN — 005, .15 ,,,,; CAUSE: Excessive launch phase dynamic loads cause Pitch IRIG to output erroneous signals to the Stab loop. RESULT: IMU drifts in Pitch axis at rate of .15 degree per second. FDO will send RTC 12 with G&N "no go." MR 4-8B. CUES: G0521 M0683 AGCU and IMU Pitch attitudes differ BACKUP: GN-007 AGC Pwr Supply Fail Sy 9 ADRK ASAP
(b) 00:03:00 PPS — 416 Sy 1 ADRK	S-IVB Main Low Valve Fails Closed Before Engine Ignition CAUSE: The mainstage solenoid fails in a deenergized state. RESULT: S-IVB will fail to ignite when the Early Staging Command is sent. FDO will command Abort. MR4-5A. CUES: G0305 and M1405. Parameters G3-401 and D1-401. BACKUP: PPS—412, Lox Prevalve Fails Closed Sy 3, ADRK ASAP. PPS—413 Fuel Prevalve Fails Closed Sy 4, ADRK ASAP.
(c) 00:05:00 PPS — 206 Sy 2 ADRK	S—II Fuel Ullage Pressure Leak SYSFLTCC, PPS — 206,, 20.0 ,,,,; CAUSE: Fuel ullage gas (gaseous hydrogen) leak

FIGURE 1.—Sample simulation exercise script—Continued.

velops a leak, designed to exhaust the helium supply in that tank well before orbital insertion. Also, around that time, one of the thrustors needed for separation jars closed; the closed one is driven by the A system. Thus, if nothing is done, only one thrustor will be available for separation. A one-thrustor separation is probably impossible: the LGC will sense the asymmetric thrust and shut the action down. So, before separation, it will be necessary either to use the PRA to read off a program that will open

the A system thrustor, or open the crossfeed valve and thus drive B system thrustors from A system fuel, or both. Of course, there will be plenty of time to do this after arrival in orbit.

But the vehicle does not arrive in orbit. The booster cuts off after about 9 minutes. And now there are exactly 60 seconds in which to do whatever needs to be done, which includes not only dealing with this thrustor problem, but also a sequence of commands required to separate the LM from the booster

SIMULATION EXERCISE SCRIPT CONTINUATION		

1.   Exercise No.
       501/LS/005b
	2. Date
	       4-20-67
		3. Page 3 of 3 Pages

4. Fault Summary		

SEQ/SGET Fault ID Sy/Se Loc	Description
	RESULT: When fuel ullage pressure drops to point where fuel pump inlet pressure is less than 20.5 psia, all 5 engines will go out (must be between 00:03:06 and 00:07:53). BSE#1 will send Early Staging Command. MR5-12. **CUES:** G0304 and M1402. Parameters D252-219, D253-219, D13-201, D13-202, D13-203, D13-204, D13-205. **BACKUP:** PPS-202, Eng 2 Fuel Prevalve Fails Closed. Sy 5, ADRK 00:07:25 PPS-214, Eng 4 Oxid Prevalve Fails Closed Sy 6, ADRK 00:07:30 PPS-235, Eng 5 Fuel Pump Degradation Rate: 3.5%/sec Limit: 0% SYSFLTCC, PPS-235,,0.0, 0.035,,,; Sy 7, ADRK 00:07:30

FIGURE 1.—Sample simulation exercise script—Concluded.

and to start the SOS. Should the controllers open the closed thrustor during launch? It is against the mission rules to send commands during powered flight, both because of the enhanced possibility of malfunction and because anything like that might distract and confuse the controllers. After cutoff, should they try to open the A system thrustor, or open the crossfeed? The latter, we thought, would take longer but might be the better solution.

Let me add that there are more than 30 communication loops available to all controllers, and most listen to 10 or more loops at a time. If it occurs to you that the overlapping voices can get confusing—you are right.

When we wrote the script, we thought that the flight director (FLIGHT) would have to make a quick, tough decision between opening a crossfeed valve, and opening a closed thrustor, or doing both. But we were wrong. In order to open the crossfeed valve, the LGC had to be in a state called

program zero zero, (P00), and this can not be done before orbital insertion. Instead, we had presented FLIGHT with a very interesting but unanticipated question: Should he or should he not violate mission rules by sending PRA 7, which would have opened up the closed A system thrustor, during powered flight. He decided not to, with the result that he failed to achieve the best possible outcome of the situation: SOS under LGC control. Unfortunately, not only had I not anticipated this as the crucial decision, but I did not respond to it quickly enough to be able to treat it as the main point of this case in the postcase interviews. So, as you will see, most of the procedures I went through in the interviews had to do with an issue that did not really require a decision. Still, I will review the procedures briefly, as a preparation for the second and more successful case.

I interviewed FLIGHT, the flight dynamics officer (FIDO), the guidance and navigation officer (GUIDO), and the officer in charge of guidance and control systems (G&C). All interviews were essentially the same. First, we reviewed the main facts of the case, to make sure that the interviewee had them clearly in mind. Then I encouraged him to talk about the issues bearing on the choice of what to do and on the question of whether it could have been done in the 60 seconds available. The main purpose of this was to make sure that we understood the technical issues properly and, in the second case, the key technical fact of the case emerged in this portion of each interview.

Next I asked the interviewee to list each possible outcome of the decision situation. For this case, each list included three possibilities. I asked the interviewee to rank these possible outcomes in order of desirability. Then I invited him to assign a value of 1 to the most favorable, a value of 0 to the least favorable, and then to use this scale to assign a value or utility to the intermediate possibility. As it happens, in this case there were no decisions to make and no information to process probabilistically, so the interview ended with a few miscellaneous probability estimates. The second case is more sophisticated.

To give some idea of the resulting utility scales, the relevant numbers are presented in table I. As you see, there is good agreement about utilities, and perfect agreement about orderings in utility.

TABLE I.—*Possible Outcomes of Case 1 and Their Utilities*

Outcome	Utility for—		
	FLIGHT	GUIDO	G&C
PRA 7 followed by LGC SOS	1	1	1
PRA 7 followed by PRA SOS	.9	.75	.8
Abort without PRA 7....	0	0	0

Those are the useful results of case 1. Let us turn to case 2. Again, I must explain technicalities so you can understand the case.

There are several kinds of aborts. Those of case 1, resulting in a suborbital sequence of tests, are called mode 3 aborts. Shortly before the whole system reaches orbit, however, it reaches a point at which, if the booster cuts off prematurely, it is possible to separate the LM from the booster, fire the LM's descending propulsion engine (DPS), and use that engine to put the LM and the remainder of the spacecraft into orbit. Once in orbit, the desired tests can, of course, be conducted leisurely, except insofar as the firing of the DPS has interfered with subsequent DPS tests because of fuel depletion. However, it is quite undesirable to attempt this unless the spacecraft actually has the capability of reaching orbit. Otherwise, it will get too far out over the Atlantic, capability for controlling it or for monitoring events on it will be lost for a while because no ground station is conveniently located, and by the time it gets in range of a station again it may be ready to reenter; thus, not even the abbreviated tests of the suborbital sequence will get done. Moreover, in this

case it may land on Africa, politically a highly undesirable event.

The main display used during launch phase is called a v–γ plot: v is the velocity of the vehicle; γ is the angle between a linear extension of the vehicle's present flight path (such as it would follow if its engines cut off and all gravitational influences were removed) and the path it will follow once in orbit. A large display of the v–γ plot is located at the front of the control room, and any controller can call up the display at any time on one of his TV tubes. A line on that plot represents the boundary between modes 3 and 4; when the dot representing the spacecraft crosses that line, it has mode 4 capability.

In order to maintain the v–γ plot, the computers must know where the spacecraft is. There are three sources of that information: radar data, collected from all over the world and interpreted by a computer complex at Cape Kennedy called the impact predictor (IP); inertial guidance calculations made on board the Saturn launch vehicle by the launch vehicle digital computer (LVDC) and telemetered to the ground; and similar inertial guidance calculations made by the LGC. One of FIDO's many functions is to decide which of these sources shall be selected to drive the v–γ plot; various kinds of reliability information bear on the selection of the source.

The point of case 2 was to destroy one source and cause the other two to disagree, with some not-very-good data bearing on which is correct, and then to initiate an abort at a time when one source says mode 3 and the other says mode 4. Then FIDO must decide which to believe, with much depending on the decision.

The first step in this process was to initiate a small progressive disagreement between the two computers. At 6 minutes after liftoff, an accelerometer in the Saturn booster was given a bias such that it reported less acceleration along the flight path than actually occurred; the effect was initially undetectable, but gradually accumulated. At 6 minutes 30 seconds after liftoff, all radar

transponders failed, causing complete loss of IP data. Then, at exactly the moment at which the LGC says mode 4 and the LVDC says mode 3, the booster was to be cut off prematurely. The only information available was the 30 seconds of information beginning when the two computers started to disagree and ending when the transponders failed.

As the case worked out, the 30 seconds of information was never used. When the discrepancy was noticed, about 8 minutes after liftoff, FIDO simply trusted the LVDC better than the LGC, and so used it as the selected source. Unfortunately, the simulation staff was a few seconds late getting the booster cutoff in, so the LVDC source and the LGC source both said mode 4, and a successful mode 4 abort was executed.

This discouraging experience stimulated us to try the case again, with a few changes to prevent the controllers from noticing that they had seen this case before. One irrelevant change was that the radar data this time failed because of a computer malfunction on the ground. Another very relevant change was that this time the booster cutoff was at exactly the right time. The third change, which was intended to be irrelevant but actually turned out to be crucial, was that early in the launch phase a leak was inserted in the gas tank that provides gas under pressure to drive the gyroscopes of the Saturn stable platform. Actually, this leak had no prospect of affecting the stable platform during the launch phase. However, when the discrepancy between the two computers developed, FIDO remembered that the accelerometers for the LVDC are mounted on the Saturn stable platform, took the stable platform pressure leak into account, concluded that because of the leak the LVDC information was less likely to be right than the LGC information, used the LGC data to drive the v–γ plot, and thus correctly called mode 4 when the booster cutoff came. This is a fine example of being right for the wrong reasons.

In interviews after the case, FLIGHT, FIDO and GUIDO all agreed that normally the LVDC would be preferable to the LGC,

simply because the LGC, as part of the LM, is new and untried, while the LVDC is a familiar system. They also agreed that the platform pressure problem was the reason for choosing the LGC over the LVDC in this case, and considered the choice appropriate.

This was a situation in which a real decision was made, in the sense that FIDO and, on FIDO's recommendation, FLIGHT, in effect had to decide between mode 3 for sure and a gamble that could lead to either a successful or an unsuccessful mode 4; an unsuccessful mode 4 is clearly less favorable than a successful mode 3. Moreover, in this case, real information processing occurred. Not, to be sure, the information processing of 30 seconds worth of old radar data that we had had in mind when we designed the case; rather, the interpretation of the platform leak. So I applied the apparatus of decision theory to the analysis of the situation.

As before, after the unconstrained part of the interview was over, I asked each interviewee (I interviewed only one man at a time) to list, rank, and then judge the value of each possible outcome. FLIGHT listed eight possibilities; FIDO and GUIDO each listed four. Table II shows the utilities assigned. The discrepancy between FLIGHT's

TABLE II.—*Possible Outcomes of Case 2 and Their Utilties*

Outcome	Utility for—		
	FLIGHT	FIDO	GUIDO
Successful mode 4	1	1	1
Try mode 4, get to Australia, refire DPS ..	.95	....	....
Successful mode 3, LGC control	.5	.75	.75
Successful mode 3, PRA control	.45	.50	.625
Try mode 4, discover cannot make orbit, do whatever tests are possible..	.3	0	0
Try mode 4, overfly Africa	.1	....	....
Try mode 4, get African impact	0	....	....
Total loss because excessive delay leads to loss of signal	0	....	....

and the others' lists is less substantial than it appears; all possibilities listed only by him have very low probability, and the last three may have been lumped in with the fourth one by FIDO and GUIDO. However, if a transformation on FLIGHT's utility judgments is made so that the fifth outcome is taken as having zero utility, then the numerical discrepancy between his judgments and those of the others is increased. But, as usual, the ordinal agreement is perfect.

I asked each interviewee to estimate some odds. First, I asked him to imagine that he was being interviewed just before the real 204L mission. Suppose, I said, that he knew for sure that he was going to lose, say, LGC data, and that the IP and the LVDC were going to disagree about the location of the vehicle. Which would be more likely to be right and, in a ratio sense, how much more likely? Similarly, for IP versus LGC in the absence of LVDC data, and for LGC versus LVDC in the absence of IP data. These three odds are related such that any two specify the third. To see this, note that

$$\frac{P(LGC\ right)}{P(IP\ right)} \cdot \frac{P(IP\ right)}{P(LVDC\ right)}$$

$$= \frac{P(LGC\ right)}{P(LVDC\ right)}$$

Naturally, the interviewees did not have this internal consistency rule in mind, and all three initially made inconsistent estimates. I pointed out the rule and the inconsistency to them and invited revision of any or all estimates; the resulting revisions were typically in the direction of greater consistency, but not enough to produce perfect consistency. For the data, see the first three rows of table III. The estimates, after revision for consistency, are sufficiently consistent. Moreover, they agree qualitatively, although, of course, not numerically.

Then I explained what a likelihood ratio is and asked for an estimate of the ratio of the probability that the stable platform pressure problem would have occurred if the LGC were going to turn out correct to the probability of that symptom if the LVDC were going to turn out correct. Communicating

TABLE III.—*Odds and Likelihood Ratio Estimates, Case 2*

Quantity estimated	FLIGHT		FIDO		GUIDO	
	1st estimate	Revised estimate	1st estimate	Revised estimate	1st estimate	Revised estimate
Prior odds for—						
LGC right: IP right..................	1:10	1:20	1:11	....	1:8	1:6
IP right: LVDC right.................	10:1	....	10:1	5:1	4:1	3:1
LGC right: LVDC right..............	1:6	1:2	1:3	....	1:6	1:2
Likelihood ratio:						
$\dfrac{P \text{ (platform problem/LGC right)}}{P \text{ (platform problem/LVDC right)}}$	2:1	....	3:1	5:1	10:1	....
Posterior odds estimated:						
LGC right: LVDC right................	2:1	....	5:1	4:1	6:1	....
Posterior odds calculated from revised prior odds and likelihood ratio:						
LGC right; LVDC right................	1:1	....	5:3	....	5:1	....

the idea of a likelihood ratio was quite difficult.

Finally, I asked for an estimate of the odds, after observing the platform symptom, of LGC being right versus LVDC being right. Having obtained that number, I pointed out that, again, an internal consistency relationship applies to the three estimates having to do with LGC and LVDC: this relationship, of course, is Bayes' theorem. In symbols, it says that—

$$\frac{P \text{ (LGC right before datum)}}{P \text{ (LVDC right before datum)}}$$

$$\bullet \; \frac{P \text{ (Datum given LGC right)}}{P \text{ (Datum given LVDC right)}}$$

$$= \frac{P \text{ (LGC right after datum)}}{P \text{ (LVDC right after datum)}}$$

The datum, of course, is the Saturn stable platform pressure problem. Finally, I invited further revision of any estimate to improve consistency; all but FLIGHT made such revisions. All posterior odds estimated are more extreme in favor of the LGC than the calculation from prior odds and likelihood ratio would imply. I speculate that the reason for this is that the case was being discussed after the truth (that the LGC was right) was known; these posterior estimates, but not the prior estimates and not the likelihood ratios, reflect this knowledge of the

truth. But in no case is the after-revision violation of Bayes' theorem large.

Now, consider the actual decision to try mode 4. For FLIGHT, if we take the value of a successful mode 3 as 0.5 and that of an unsuccessful mode 4 as 0.3, we see that for any odds for success of mode 4 greater than 2 to 5 he should choose mode 4; on any estimate, his odds for successful mode 4 were greater than that. Similarly, for both FIDO and GUIDO, if we take the value of a successful mode 3 as 0.75, we find that any odds for success greater than 3 to 1 should lead to trying for mode 4; their estimated odds were greater than 3 to 1. The calculated 5-to-3 odds for FIDO raise a problem; the value of a successful mode 3 would have to be less than 0.625 to justify trying mode 4 with those odds. That happens to be midway between his two value estimates for mode 3; but I assume he considers mode 3 under LGC control much more likely in this case than the less desirable mode 3 under PRA control. Still, with that minor exception, the decision looks entirely reasonable in the light of the value and probability estimates.

That ends the data analysis. What does it all mean?

First, I should caution you that my conclusions grow much more from the information I picked up while performing this study

than from the outcome of the study itself. The study, as, of course, you have long since recognized, is clinical rather than experimental in nature. As with all clinical procedures I know of, it is suggestive to intuition, but not suited to proving what it suggests.

A first conclusion is that nothing in this exercise denies, and much confirms, the major premise with which I began: that every decision, including those made on line by space vehicle controllers, depends on subjective answers to the questions what is at stake and what are the odds. And the results also support an unstated but clearly implied premise: that decisionmakers can report fairly clearly and precisely on their subjective stakes and odds if you ask them properly. Incidentally, I should emphasize that I did not ask them properly, particularly about odds and likelihood ratios. To do so, I should have first given them 5 hours or so of training in how to estimate these quantities. Had I done so, I am confident that the residual incoherences in the data would have vanished. As it is, I feel they were extraordinarily consistent, considering that they had only 10 minutes of explanation of some of the most difficult concepts of modern statistics before being asked to use them, not arithmetically, but intuitively.

A second conclusion is that irreducible uncertainty plays a remarkably small role in on-line control of space flight. This conclusion will probably horrify controllers who hear it; they must feel that they struggle with uncertainty all the time. And of course they do. But what they struggle to do is to reduce or eliminate the uncertainty, and mostly they can do so very well. The reason is simply that the capability of eliminating uncertainty is one of the major design goals of the entire system. Rare, indeed, is the question about which three or four or five definitive sources of information are not designed into the system; in order to set up our cases, we had to produce quite artificial combinations of malfunctions to frustrate the purpose of this redundancy of information sources. Such combinations can and do occur in real space flight, but they are very rare.

A third conclusion is that value judgments play an extremely major role in space flight. Very nearly every decision that must be made, ahead of time or during a mission, requires juggling the various goals of the mission, choosing how much to compromise one goal in order to achieve another; of course, this is especially true in case of a serious malfunction. So it is in what it has to say about values, not what it has to say about probabilities, that decision theory is most relevant to mission control.

A fourth conclusion, growing out of the last two, is that PIP, an on-line system for diagnosis that I have been advocating in many contexts recently, has little, if any, prospect of being useful in on-line mission control. The reason is simple. PIP is a method for combining Bayes' theorem with human judgments to perform diagnostic information processing. It requires a preselected exhaustive set of mutually exclusive diagnoses, and this set cannot be too large. Moreover, it is best suited to exploiting inconclusive data. But in on-line diagnosis of spacecraft malfunction, the set of possibilities is extremely large and initially unspecified, and the data are in general conclusive, in the sense of conclusively eliminating hypotheses.

I hasten to add that the basic ideas of PIP are, nevertheless, very relevant to mission control; I will explain in a moment.

A fifth conclusion is that the key place to apply decision theory, including both its diagnostic aspects (and diagnostic tools like PIP) and its value judgment aspects, in mission control is in writing the mission rules document. The mission rules document is a large book of premade decisions, an attempt to anticipate all possible contingencies and decide what to do if each occurs. It now uses an explicit but ordinal set of value judgments and no explicit but many implicit probability judgments. I believe that addition of cardinal values, explicit probabilities, and rules for combining and using them to the armament of those who write mission rules would be extremely helpful, particularly in resolving difficult cases—which, I am

told, are frequently encountered in writing the document. Especially important in thus formalizing the development of mission rules would be a mechanism for combining inconsistent value dimensions; the weighted linear average rule so prominent in formal decision theory is the obvious candidate.

I would like to conclude with a question and an observation. The question is this. Immense amounts of money, engineering effort, and complexity have been invested in protecting controllers and astronauts against uncertainty. Was it worth it? If information systems (as distinguished from propulsion, life support, and other systems that do something more than just transmitting and displaying information) were less redundant and less oriented toward certainty, the whole system would be less complex. Quite possibly, the gain in reliability resulting from decreased complexity would exceed the loss in reliability resulting from reduced information availability. Even if it did not, the loss in reliability might be minor compared with the savings in money. I do not know the answer to my question; I wonder if it has been carefully asked.

My concluding observation has to do with the future of mission control and mission controllers. It assumes that manned space flight has a post-Apollo future, an assumption with which I hope Congress comes to agree. It also assumes that the future will consist primarily of interplanetary flights, necessarily lasting for weeks or months.

The original Project Mercury controllers were highly trained engineers and highly motivated, intensely committed men. According to my informants, both the level of

training and the level of motivation is now lower than it was in Mercury days, though still very high. Most control procedures depend on the high quality of the controllers, and would be impossible otherwise.

But really long missions are going to consist mostly of periods during which nothing is happening. Even if such able controllers are available, they are not going to be willing to spend weeks or months of boring, inactive console duty. Yet there will often not be time, if an emergency arises, to call in the first team of controllers from their other activities. The implication, it seems to me, is that lower level personnel, such as are now found in FAA Air Traffic Control Centers, will necessarily come to man the NASA consoles also. They will be high school, not college, graduates. Their work will be to them a job, not a priesthood.

Present procedures, or extensions of them, just will not work if such men are manning the consoles. The computers will have to do much more of the job. Procedures will have to be far more formalized and prespecified than they are now. Detailed technical knowledge of all the systems, spaceborne and ground, cannot be assumed.

In such an environment, formal decision theory is very likely to be crucial to making the on-line decisions. The form this will probably take is that the mission rules will be oriented toward the use of decision theory, and the computer will know and apply them to new situations as they arise. If this is the shape of the future, then NASA should begin now to find out how to move toward it. Applying formal rules to the writing of mission rules documents is a good first step.

APPENDIX A

On September 17, 1968, two more launch abort cases designed to present controllers with decision problems were run, and I interviewed controllers as before. These simulations were for the Apollo 7 mission, whose objectives and personnel were different from the LM-1 mission used for the studies reported above. The Flight Director, Glynn Lunney, was fully as helpful as Kranz had

been. During the simulations, astronauts in a simulator at Cape Kennedy were in voice communication with the controllers; the instruments in their simulator, linked with the simulation computers, behaved as though they were in flight. But the launch phases of the two missions are very similar, and the role of the astronauts in these two cases was entirely peripheral (except during debriefings).

In the first case, a leak was started at launch in the service propulsion system (SPS) engine oxidizer supply. (The SPS is the main thrust engine of the command service module (CSM).) The instrumentation prevents determination of which is leaking—the oxidizer or the helium contained in the system to force the oxidizer into the engines. Then the booster failed in the mode 4 region. Normal procedure, in the absence of the leak, would be to separate the CSM from the booster and get into orbit by firing the SPS. If the leak is only a helium leak, this would still be the best procedure. But, if the SPS is fired posigrade (that is, in a direction intended to produce orbit; the opposite direction of firing is called retrograde) and cuts off before orbit is reached, as it will if the oxidizer has been badly depleted, the resulting impact point may be in some very undesirable location, like Africa. Moreover, even if orbit is reached, there may not be enough SPS oxidizer to permit deorbit; in that case, the difficult and hazardous procedure of reentering using the reaction control system (RCS) thrustors may have to be used. Moreover, the loss of large quantities of SPS oxidizer affects the SPS burn in two other ways: the changed mass of the vehicle will have to be allowed for, as will the possibility that the changed center of gravity of the vehicle may not be adequately compensated for in the thrust vector control (TVC), which controls the exact direction of firing of the SPS loop.

This case was designed to highlight a loophole in the mission rules, which fail to specify what to do when it is uncertain whether the leak is of helium or of oxidizer. Linking a case with a flaw in mission rules both highlighted the flaw and gave some feeling for whether explicit value and odds judgments could be helpful in formulating mission rules.

In the simulation, the leak was discovered at about 1 minute 44 seconds after liftoff and was carefully tracked. After the booster cutoff, a controller recommended to the Flight Director that they do an SPS burn to orbit. But the Flight Director decided not

to burn the SPS. Instead, they successfully executed a mode 2 abort.

In debriefing, the Flight Director explained that G&C had reported that if the leak was an oxidizer leak, then the impact point would have been somewhere in Africa; that is why he chose to do the mode 2 abort. If he had been committed to a landing regardless of which abort he tried, he would have chosen the SPS burn to orbit—the mode 4 abort.

In the interview afterward, the Flight Director clarified that the leak rate was quite slow for a helium leak, but of reasonable rate for an oxidizer leak; this evidence favored the hypothesis that it was an oxidizer leak. His decision, however, was based more on value than on odds considerations. His estimate of 1 to 1 prior odds may have reflected a feeling that he had initially no relevant information rather than a judgment that the two kinds of leaks are equally frequent; later informal comments suggest this. His value and odds estimates are presented in table A–1. Given these, he clearly did the right thing.

TABLE A–1.—*Values, Odds, and Likelihood Ratios, Case 3*

Quantity	Flight Director	G&C
Estimated utility for—		
Successful mode 4.......	1	1
Successful mode 2.......	.90	.50
African impact	0	0
Prior odds: liquid to gas leak	1:1	1:19
Likelihood ratio	3:1	7:1
Posterior odds	3:1	3:1
Calculated posterior odds ...	3:1	7:19 (=1:2.7)

The other interviewee in this case was G&C. His reasoning was qualitatively like that of the Flight Director; his estimates are also contained in table A–1. When the inconsistency between his prior odds, likelihood ratio, and posterior odds estimates was pointed out to him, he felt uncomfortable about it, but did not wish to change any of

the three numbers. (Time was too short to permit careful exploration of his opinions.) In spite of the substantial difference in values between G&C and the Flight Director, the odds and values of each would lead to the decision actually made.

The final case was once more a launch abort, this time in what is called an apogee kick situation. This is a case in which the booster has burned long enough to produce orbit, but the perigee of the orbit is too low. In this case, further propulsion is necessary to produce a satisfactory orbit, but a substantial delay in initiation of that propulsion is desirable.

There are two alternating current (ac) buses. Normally, current from ac bus no. 2 operates the thrust-vector-control (TVC) system for the SPS. The first malfunction in this case caused the astronauts to lose all capability of monitoring ac bus voltages. The second malfunction was a short in ac bus no. 1, which the controllers were expected to diagnose without too much difficulty. The final and crucial malfunction was an intermittent partial short on ac bus no. 2 that causes its voltage to oscillate around a marginal value. Then, when premature booster cutoff made an SPS burn necessary, the Flight Director had to decide whether to go ahead and call for the burn, with the possibility that lack of TVC would cause the spacecraft to tumble, or whether to initiate an RCS deorbit, with probable landing in the Indian Ocean. The purpose of the case was to highlight the arbitrary character of mission rules built around "magic number" cutoff values.

As the case worked out, all malfunctions were detected and treated as expected. When the booster cutoff came, it was treated as a mode 4 case, not an apogee kick case. (These cases are actually different only in the time of initiation of the SPS burn; little harm results from treating an apogee kick case as a mode 4 case.) The SPS burn was called for and executed. In debriefing, the Flight Director said he had assumed the TVC system was satisfactory.

In an interview afterward, the Flight Director explained that the crucial point is that all systems driven by bus no. 2 were working satisfactorily, regardless of the marginal and fluctuating voltage readings. The electrical environmental and communications systems engineer (EECOM), interviewed separately, agreed.

The value judgments and probability judgments were both interfered with, in both interviews, by shortage of time. Both men agreed that a successful mode 4 was best, and that an unsuccessful mode 4 leading either to a land impact or a tumbling spacecraft was worst. The other possibility, which both considered, was that the mode 4 might get the spacecraft into orbit, but that the buses might thereafter both fail, making immediate deorbit necessary. The Flight Director valued that at 0.7 to 0.8; EECOM valued it at 0.6. Both agreed that the odds were high in favor of ac bus no. 2 being able to operate the TVC system. The Flight Director put those odds at 30 or 40 to 1; EECOM put them at a million to 1. For either figure, the decision made was obviously correct. I believe these numbers would all have been somewhat different (especially EECOM's) if more time had been available.

The conclusions from these two cases are basically the same as the conclusions presented in the body of the presentation. Both cases were designed to exhibit how explicit value and probability judgments can be used to guide the selection of (necessarily arbitrary) cutoff numbers for use in mission rules. In the first case, the cutoff is on a probability judgment; in the second, it is on a voltage. One byproduct was the observation that information on which such cutoffs should be based (e.g., What is the minimum voltage required to operate the TVC system, and what will happen below that voltage?) may not be easily available. The manufacturers of equipment provide such numbers—usually with a built-in safety factor. When such safety factors are piled on top of one another as more and more systems interact with one another, they may become entirely unrealistic. But performance testing in the spacecraft under actual mission conditions

is impossible, and performance testing on the ground is both costly and of doubtful relevance. Moreover, exhaustive exploration of tolerances of all systems to all possible deviations from design parameters is preposterous; it would take too long. Clearly what is needed is more explicit tolerance-of-variability information from the manufacturer—without any built-in safety factors. Such numbers would provide the probabilities that, combined with the value judgments for the various possible outcomes of the actions being considered, would permit the optimal selection of cutoffs in those mission rules that must be based on such cutoffs—as many must.

Unfortunately, simulations cannot substitute for work with the real systems in selecting these cutoffs. The simulation must simply accept the design information provided by the manufacturer; it cannot examine the precision of that information.

GLOSSARY

AGS—Abort guidance section: An auxiliary guidance system on the LM. (See **PNGS**.)

AOS—Acquisition of signal: The moment at which a ground site is able to receive signals from and transmit them to the spacecraft. (See **LOS**.)

apogee kick—When booster has burned long enough to produce orbit, but the orbit perigee is too low. Further propulsion is necessary to achieve satisfactory orbit.

APS—Ascent propulsion system of the LM. (See **DPS**.)

A System—One of the two fuel-and-oxidizer systems serving the RCS thrustors. (See **B system** and **RCS**.)

ATC—Air traffic control.

Booster—The controller who monitors Saturn launch vehicle engines.

B System—One of the two fuel-and-oxidizer systems serving the RCS thrustors. (See **A system** and **RCS**.)

CAP COM—Capsule communicator.

COI—Contingency orbit insertion: What is done in a successful mode 4 abort.

CSM—Command service module.

DPS—Descent propulsion system of the LM: Pronounced *dips*. (See **APS**.)

EECOM—Electrical, environmental, and communications systems engineer: A controller.

EMR—Engine mixture ratio.

FIDO—Flight dynamics officer: A controller.

G&N; GNC—Guidance and navigation systems controller.

GUIDO—Guidance officer: A controller; not the same as G&N or GNC.

IGM—Iterative guidance mode.

IGN—Ignition.

IMU—Inertial measurement unit.

IP—Impact predictor: A computer complex at Cape Kennedy that processes radar data.

IRIG—Inertial reference integrating gyro.

IU—Instrumentation unit of the SLV, containing the LVDC.

J2—An engine on the SLV.

LGC—LM guidance computer.

LM—Lunar excursion module (formerly LEM).

LMP—LM mission programer: A group of programs in the LGC.

LOS—Loss of signal. (See **AOS**.)

LVDC—Launch vehicle digital computer, located in the IU.

MAP—Message acceptance pulse: A signal that the spacecraft has accepted a command.

MCC—Mission Control Center, located at MSC.

MED—Manual electric device.

M&O—Maintenance and operation.

MOCR—Mission operations control room, part of the MCC; controllers sit here.

Mode 3—An abort in which COI is not possible.

Mode 4—An abort in which COI can be achieved by separating the LM from the SLV and firing the DPS.

MSC—Manned Spacecraft Center (Houston, Tex.).

PLUS X—Forward direction, parallel to spacecraft.

PNGS—Primary navigation and guidance system of the LM; pronounced *pings*.

Poo—Program zero zero: A null program in the LGC.

posigrade—In the direction of orbit.

PRA—Program reader assembly: A magnetic tape reader in the LM.

RETRO—Retrofire officer: A controller.

retrograde—Opposite to the usual orbit direction.

RCS—Reaction control system: A system of small attitude-control rockets on the LM.

RTC—Real-time computer.

SIM SUP—Simulation supervisor: Controls the simulation staff during simulations.

SLA—Spacecraft-LM adapter: The unit that connects the Saturn to the LM.

SLV—Saturn launch vehicle.

SOS—Suborbital sequence, executed in case of a mode 3 abort.

SPS—Service propulsion system.

TCA—Thrust chamber assembly.

TVC—Thrust vector control.

ullage—Empty space in a fuel tank. To make fuel flow out of a tank, ullage must be produced; this can be difficult under 0-g conditions.

v-γ—Plot of spacecraft velocity against flight path angle.

Discussion[1]

EDWARD M. HUFF

Ames Research Center

EDWARD M. HUFF: The suggestion has been made that we deviate from the schedule at this point and try to elicit as many responses as possible concerning the two papers. I will just summarize briefly. I do not know that I can say anything that has not been said or implied already.

I think Miller's paper gives us an excellent general summary of flight-control operations. I think, by implication, that we can see various ways in which academic psychology can fit into the picture. Whether we want to refer to all the potential applications as decisionmaking is a subject I will not go into. I understand that Rathert has some comments concerning relationships between Edwards' discussion and the SST control operations, so I would like to turn the conversation over to him.

GEORGE A. RATHERT: The point occurred to me that Edwards' remarks were, to an extent, forecasting future trends of manned-control-center requirements. I am struck by the similarities between simulation complexes for overall mission control for the SST and simulation complexes for the overall mission control for post-Apollo. The manned control center is being dragged down to the level of air-traffic control (ATC), because of the degradation of personnel quality. On the other hand, the SST with its multiple redundant systems, multiple potential for failure, is approaching the complexity of the

spacecraft that Houston control works with. I do not know whether manned space operations and operations of the SST are heading toward the same point, whether they are still several miles apart, or whether there has been crossover. I do know that there is great urgency in solving the problems connected with both. You cannot afford to make a mistake. In both of these cases, you will have to resort to decisionmaking technology, the academic psychological impact that Edwards is talking about. I think that both operations will become similar enough so that the people responsible for these operations will actually listen to people like Ward and let him participate directly in operational applications. I do not know where these trends will start, but a group like this that knows both ends of the spectrum is probably a pretty good place. I think this point should not be missed.

Edwards' type of analysis and observation at Houston is equally applicable to the struggle we see going on in the FAA over the ATC systems. There are several differences that are interesting. The command control center in the SST will be in the cockpit. In the example that I used—flying from London to New York and having a combination of in-flight systems failures making it necessary to divert and land at Gander—the pilot on the flight deck makes the decision. Then he notifies ATC and uses its facilities to get himself a clear air path and approach. I think, perhaps in the long-range missions at Houston the same thing will happen. Im-

---

[1] The discussion concerns Miller's and Edwards' papers.

107

mediate command will move to the vehicle itself. Communication problems and a lot of things point toward this, but the application remains the same. Aids in decisionmaking remain the same.

Suppose, for instance, that the pilot of an SST breaks out of clouds and has to judge his landing approach. This is something every pilot does in every airplane. But, in the case of the SST, the pilot may be sitting there saying, "I have got a triply redundant longitudinal control system. One control has failed. Of the two remaining, there is a certain increased probability that they will fail and compromise my ability to make a successful landing." In this case, the decision should be made on the basis of weighing factors that are based on a statistical history of those control systems in that particular configuration. The pilot is not going to carry this information in his head. It must be stored, evaluated, and presented to him in a form that he can use.

HUFF: In this particular case, as you envision it, you should have a well-trained specialist making the ultimate corrective actions?

RATHERT: I am not really deeply involved in this area. When I go to Boeing and say, "I am going to simulate an overall mission control, what is it going to be?" The answer I get does not give me any confidence that they know of Edwards' work.

STEVEN E. BELSLEY: I do not necessarily disagree, but I believe that there is an application of decisionmaking that is of more importance than the traffic-control problem. There is a need to establish minimums under which a pilot will land an aircraft. This is a real-life situation. I read in *Aviation Week* that some simulator tests were run by the Air Line Pilots Association (ALPA). Now I understand that the ALPA, the airlines, and the FAA are all in a three-ring circus arguing with one another. The FAA has already taken a position. The manufacturers are happy. The pilots are mad. You ask them if they have any data on which they base the setting up of these rules, which is really a decision matrix, and you cannot find out.

This is a very important question. It is already with us in the present jet transports, and it is going to be with us when the bigger jets have 300 or 400 people aboard.

WARD EDWARDS: First, let me say that decision theory is no panacea for decisionmaking. Fisher once said about analysis of variance that it was just an orderly way of arranging the work, implying that the computations of means and variances were what people were going to do anyhow. In just that same sense, decision theory is really little more than an orderly way of arranging the process of making decisions. It will not end arguments between persons with different points of view or different value systems. Instead, it will focus the arguments by inviting people to consider the detailed issues about which they are arguing and by providing a language and a set of numbers to embody the result of that consideration.

Second, Mission Rule Documents are only an example of a very large class of documents that constitute elaborate sets of premade decisions. You just named another. I have in my plane a briefcase full of approach plates generated by Jeppersen & Co. Each such approach plate contains a table of the ceiling and visibility minimums for instrument approaches. Of course, as you rightly point out, every such table is a set of premade decisions. I am utterly confident that the decisions were made by an unthinking application of a set of formal rules. Some thought probably went into the rules, but I am quite sure none went into the application in many of the individual plates. Of course there are hundreds of other examples of such precanned decisions. I think it is quite likely that formal structuring procedures can be applied to every instance of writing such documents, where it is socially worthwhile to spend the additional effort, time, and money to do so.

JOSEPH MARKOWITZ: In one of the cases you brought up, there was a disagreement in redundant systems. The trend in instrumentation displays is to go to displays that may be predictive but, in a sense, contrary to the information that one needs to make decisions

in divergent redundant elements. What you would like to know is the past history over some point in time. You currently just allow whatever memory the operator has. I am suggesting that in the SST program, where you do not have elaborate support personnel, there is an advantage in having instruments that will carry some past history with them.

RATHERT: The jet engines are presently carrying instrumentation for measuring certain critical parameters that indicate the need for an overhaul. In fact, instruments are continually monitoring the specific engine to predict its failing point so that it can be overhauled just short of that point.

BELSLEY: There is another point I would like to make. A study, run by Serendipity Associates, indicated the desirability of having a flight director aboard the SST-type aircraft. The flight director would not be a flyer but would command the system as the captain does from the deck of a ship. This type of a thought has to be examined, because you can predict on the basis of workload analyses and timeline analyses that there are certain parts of the flight where three members of the SST crew would reasonably be overloaded.

JEROME I. ELKIND: How did Serendipity go about predicting that the workload was going to be so high that another man would be needed?

BELSLEY: They figured on a time-available basis that each man had just so much time for each job. They plotted every minute.

RATHERT: It was a statement of all the functions required on board.

ELKIND: Has anyone gone through that kind of an analysis for the mission control center?

HAROLD G. MILLER: Yes. Every time a task occurs that we cannot perform with the people we have.

RATHERT: I think the need for a better approach to this is quite obvious from the current controversy over the 737, two pilots versus three. The argument is going on between the head of a pilots' union and the certification authority, neither of whom know anything about the human-factors problem involved. They have no professional help.

BELSLEY: You mean that there is apparently none. There may be some, but we do not know about it.

Examining the problem is very simple. The airline operator wants to have a minimum amount of people on board, because of salaries. The Air Line Pilots Association likes to have as many people as possible; they are also concerned about safety. This is what must be decided in some manner, shape, or form. I am not aware of how one goes about making this determination.

MARKOWITZ: In this case, where there was a decision to go PRA–7 at some particular point, the decision had been made. Is there a way the decision could be delegated from flight so that it could be made automatically at the appropriate time?

MILLER: They talked about what they could do automatically to get to the SOS program. They say, "If this situation happens, rather than going through the flight director, go ahead and implement." That is always dangerous. And probably the man that made the decision to go PRA–7 was the man on the front row, G&C.

STANLEY DEUTSCH: Instead of asking the various members of the flight-control crew what their estimates are in terms of "Should we take mode A or mode B?" why not punch these estimates in the computer and have it provide guidance for an answer.

MILLER: I think that this is what Ward was getting back to. There are so many combinations and possibilities of things that if you took a specific case, I think you could do it; but the probability of ever using a specific case may be fairly small.

EDWARDS: I can tell you how to set up a program for any 10 possible malfunctions, perhaps even for 50. But you are dealing with numbers of possible maufunctions in the many, many thousands. Sometimes in a welter of possibilities some are more probable than others. So you can set up a procedure to take care of the probable ones and provide an escape route into the direct human judgment for the improbable ones. But that is

not the situation at MSC. Here they are all improbable.

DEUTSCH: The kind of problems you presented earlier struck me as having very logical answers. Now you are getting down to value judgments on the part of the people involved, GUIDO, FIDO, G&C, and so forth. I have the feeling these judgments could be preprogramed and still not solve the problem.

DOUWE YNTEMA: If I could make a comment. . . . In a few years it will be entirely practical to have the mission control books on-line and accessible much faster than a human being can turn the pages.

EDWARDS: I assume that can be done. But the mission rules are already condensed representations of an elaborate decision process that would be elaborate whether it were done in the head or in some more formal sense.

What will happen is that the decisions will be premade and stored in the computer. However, I do not think that the decision process will go on-line in the computer.

DEUTSCH: It strikes me that the advantage of the equipment is that it provides the routine work for the man, leaving him to make the final decisions. I am aware that these are random emergencies and perhaps the probabilities are very low, they are multiplicative and, over a period of time, some of them will occur. But if the pilot can ask certain questions of the computer and the computer can provide answers, then the pilot still must make the final decisions. I do not think that you will ever take the pilot out of the decisionmaking loop. But he now has obtained routine information that he can use to make decisions, instead of having to ask everyone else for what should be standardized information.

Signal Detection[1]

JOSEPH MARKOWITZ

Bolt, Beranek & Newman, Inc.

I find it fitting that the topic of signal detection should be included within the framework of a conference designed to bring theoretical advances of a portion of science closer to the applications demanded by men in the field. I find it fitting because, of the topics listed here today, signal detection has most consistently, even flagrantly, flirted with the realm of applications.

For the moment, let me broadly define "signal detection by the human observer" as the processing of any sensory input. And let me furnish you with a capsule version of the historical forerunners of current research in signal detection.

Man's concern with his own sensory apparatus began quite naturally with an interest in causality; that is, what aspects of man's sensory system, or what elements of the physical world around him, combine to produce the sensations as he knows them? It became clear, however, that certain aspects of sensory functioning were more interesting at the quantitative level as opposed to the qualitative level. The field known as psychophysics is an outgrowth of such quantitative endeavors.

There was, for example, the work of Weber. His concern was to ascertain what minimum increment or decrement in the energy of a physical stimulus would produce a noticeable change in sensation.

In Weber's work, we have the first of a number of examples that show how close psychophysics has been to the applications so often demanded. Indeed, what could be more applications oriented than a table that tells us how much additional energy need be added to, or subtracted from, a stimulus to make a noticeable impression on a human observer?

In fact, however, the just-noticeable difference of Weber was destined to play the role of a building block in the theoretical developments of the father of psychophysics, Fechner. Fechner's abiding interest was in the measurement of sensation. He saw in Weber's work the unit for his scale, which could be used as a yardstick of sensation. That is, Fechner would construct a scale of sensation that used as its unit a just-noticeable difference (assuming they were everywhere equal). And so we see that an idea of Weber's, promising for its applications, turns out to be grist for the mill of a theoretician.

Nor was the just-noticeable difference the only practical concept that Fechner levitated to higher purpose. Consider the threshold of sensation: that amount of stimulus energy which, if increased fractionally, will lead always to a sensation and which, if decreased fractionally, will never provoke a sensation. Who could deny that knowing the threshold for a given sensory modality, for a given observer, would be useful and practical?

For Fechner, however, the value of the threshold was rather as the zero point in his scale of sensation. Once again a practical

notion becomes but a piece in a theoretical tinkertoy.

Nonetheless, we might applaud Fechner's attempt to measure sensation. Had he but been successful I have no doubt that each of us could have put his scale to good use. Instead we find the literature brimming with discussions of fine theoretical points rather than illustrations of the utility of his scale.

Closer to us in both time and interest is the development of signal-detectability theory—a development that runs a remarkably parallel course from practical problems to abstract considerations. Indeed, signal detectability theory is, in its way, an instrument for measuring sensation, at least for weak stimulation. We prefer to think that it is a more sophisticated measuring instrument, defining, as it does, its zero point and unit statistically rather than absolutely. Moreover, the theory recognizes that the apparent sensation of an individual is a function not only of his sensitivity to the stimulation, but of his biases of predispositions as well.

Others present today could better describe the genesis of signal-detectability theory. Its early development was intimately related to the practical problems of target acquisition and identification by radar operators. In addition to the hardware constraints of the radar system, such as antenna gain, directionality, bandwidth, and scope resolution, the human operator introduced vagaries of his own into the system. The operator was, it seems, able to trade off certain kinds of errors, such as missed targets for other types of errors, such as acquisition of targets that did not exist. Ignorance about the way in which these tradeoffs were made prevented accurate quantification of the performance of the radar system as a whole.

Signal-detectability theory was able to do several things in this context. First, it quantified the trading relationship between the various aspects of the observer's performance. Second, it was able to ascertain the observer's criterion. Third, it was able to isolate certain variables that might influence the setting of that criterion. Fourth, it provided a way of estimating the sensitivity of the total system independent of the criterion held by the observer. And, finally, it was able to set up optimal standards of performance against which the actual performance of the observer might be judged.

Further applications of the theory must have looked promising indeed. For example, it turns out that the variability associated with any measurement of the system depends heavily upon the value of the criterion that the observer employs.

Thus, the ability to manipulate that criterion could lead to increased efficiency for each measurement of the system's performance. Alternatively, the cost of an observation, perhaps in terms of the risk involved in the measurement in a field situation, depends also upon the criterion. Again, being able to manipulate this criterion should reduce the cost of measurements on the system.

Moreover, the ability to measure and correct an observer's criterion in a training context, so that it reflects realistic costs and values, should be particularly useful. Finally, the theory seems to offer the comparison, under certain conditions, of complex and dissimilar systems, or evaluation of system performance when subsystems are changed.

It turns out, however, that relatively little work was done in actually applying the tools of the theory. Instead, a great deal of work was put into refining certain aspects of the theory and establishing its scientific validity, rather than establishing its practical utility.

A portion of our work has been devoted to trying to reverse this trend and broaden the application of the theory. An example of expanding the theory to new applications is our use of certain laboratory recognition experiments to study and evaluate road signs.

The use of recognition experiments is certainly not new to investigations of this type, although the specific form of the subjects' responses that we solicited departs somewhat from tradition. In addition to asking observers to tell us which of a set of stimuli was presented on a given trial, we asked also for a rating of confidence on, say, a four-point

scale. Acquiring raw data of this type allows us to apply certain aspects of signal-detectability theory. As I mentioned previously, the chief advantage of such an analysis is that it yields a pure measure of recognizability uncontaminated by the biases or predispositions of the observers.

To give an example, the costs and values of driving on the highway may predispose individuals to give a "stop-sign" response with the slightest of provocations. That is, a driver may have a strong "stop-sign bias." As we shall see, the theory allows us to assess separately the recognizability of a stop sign and the bias toward responding stop sign. Moreover, the index of recognizability that we arrive at is a pure number and, as such, allows not only easy comparison of different signs, but allows as well the quantification of the deleterious effects of such independent variables as, for example, glare.

In brief, we have applied signal-detectability theory as follows. When the observer is presented with a stimulus, say a picture of a stop sign, we can consider his choice as a binary one between the response stop sign and the set of all other permissible response alternatives. In fact, we could put precisely this question to the observer: "Was that a stop sign you just saw? Yes or no?"

Because we could have asked this question about any one of the response alternatives, no loss of generality is implied. Moreover, we could have questioned the observer about whether a stop sign was presented when, in actuality, some other sign had been presented. In this example, then, there are only four possible combinations of events, and with repeated trials we can estimate the relative frequency, or probability, of each event and enter these in a table such as table I.

As we have indicated, all the information in the table can be summarized by the two probabilities p_1 and p_2. If an observer suddenly increased his predisposition to respond stop sign, we would expect both p_1 and p_2 to increase. We would not, however, be willing to say that the stop sign had suddenly become more recognizable. What we need to know is how the two probabilities can be

TABLE I.—*Four Possible Combinations of Events and Their Probabilities*

Stimulus	Observer's response	
	Stop sign	Other
Stop sign	p_1	$1 - p_1$
Other	p_2	$1 - p_2$

expected to change in concert with each other, for a constant level of recognizability of the sign. This is where the theory helps us.
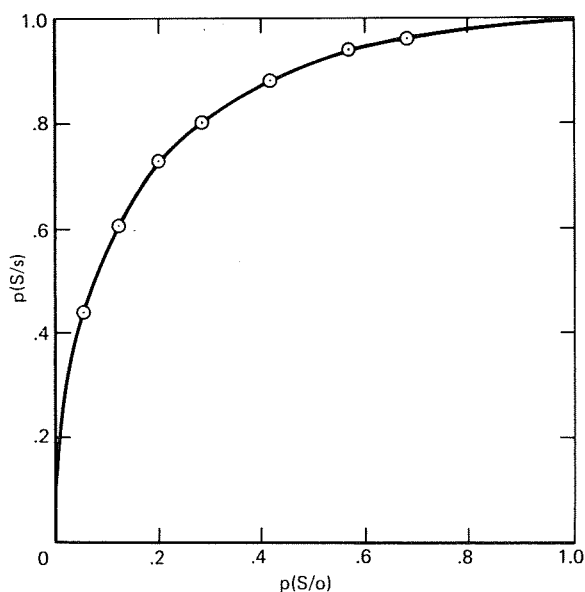

FIGURE 1.—Example of the conditional probability of the stop-sign response given the presentation of a stop sign as a function of the conditional probability of the stop-sign response given the presentation of some other stimulus.

The theory tells us that if we plot one probability against the other for various biases, we can draw a smooth curve through the points, as shown in figure 1. In this context we might call this an equirecognizability curve. If a second sign under identical conditions yielded a curve above the one shown, we could conclude that the second was more recognizable.

In short, we need to know what the curves for two stimuli look like in order to compare

them. Recent developments in signal-detectability theory allow us to obtain these curves from the ratings that, as I mentioned, we solicit from the observers.

In practice, before dealing with the complex sign, we dealt with the design elements of the sign system. Let me briefly describe an experiment conducted to determine the recognizability of shape.

We felt that in real life the chief constraint is the time that must be spent in actual viewing if one is to differentiate between alternative shapes. As a consequence, extremely brief visual exposures were used during the laboratory testing. A slide was projected for perhaps 10 to 30 thousandths of a second. The observer had then to indicate which of the alternative shapes he thought was shown. He also had to indicate numerically how sure he was.

In figure 2, we see the alternative shapes used in the experiment. Each observer had a sheet showing the alternatives.
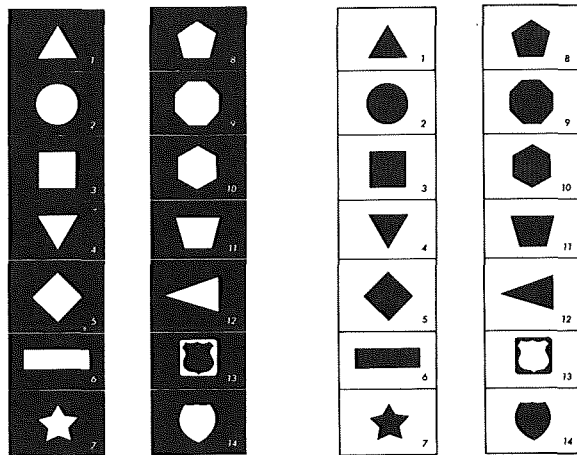

FIGURE 2.—The shapes used in the shape-recognition experiment.

Figure 3 shows the format of a portion of the data after they were analyzed according to the guidelines of signal-detectability theory. The data are for a single shape, the triangle, and represent the results of a number of observers over a number of trials.

I have changed the scale along each axis, as suggested by the theory, in order to be
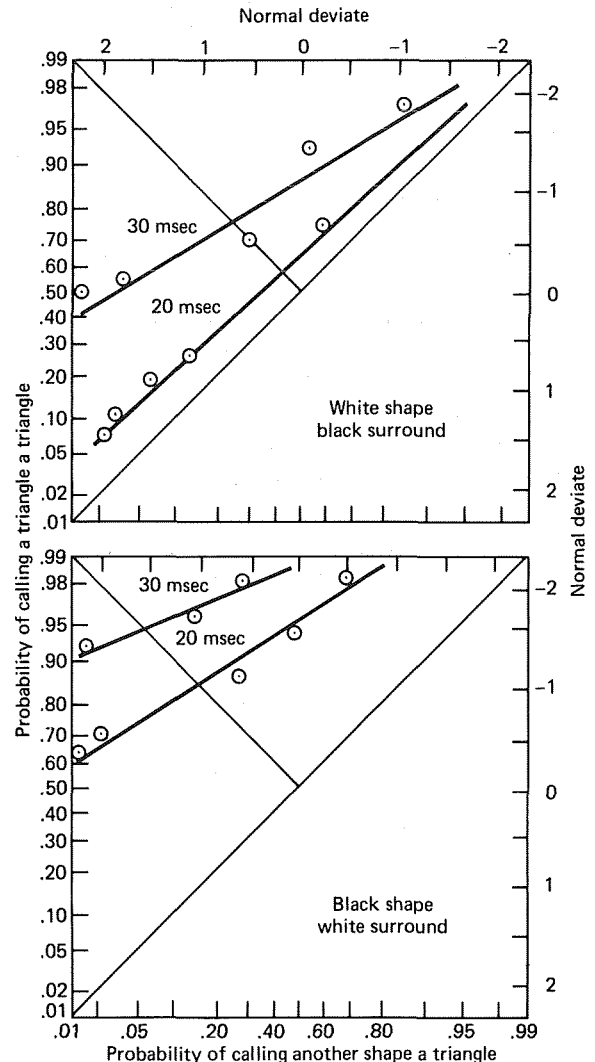

FIGURE 3.—Selected data from the shape-recognition experiment.

able to fit straight lines to the data rather than curves of the sort that you saw previously.

The vertical axis shows how often a triangle was actually called a triangle. That is, the vertical axis shows how often observers correctly identified a triangle. But, remember that this is not the whole story. We realized that an observer could increase the percentage of times he was correct in calling a triangle a triangle simply by saying "triangle" most of the time. In the extreme, he would correctly identify every triangle by

stubbornly saying "triangle" each and every time we showed him any shape. That is why we include the horizontal axis. The horizontal axis shows how often one of the other shapes was identified, mistakenly, as a triangle.

The line that best fits a set of points gives us the trading relation of these aspects of observer performance, taking into account the observer's bias toward saying "triangle." The actual degree of recognizability is given by how close to the upper left-hand corner the data fall. Thus we can see that when a triangle is presented for 30 thousandths of a second it is more recognizable than when it is shown for only 20 thousandths of a second, as we might expect. The positive diagonal represents purely chance performance.

The data on the top are for a white triangle on a black background. On the bottom are the data for similar exposure times for black triangles on a white surround, and you can see, the improvement in recognizability. As you can see, both lines move up toward the upper left-hand corner. The data were similarly treated for each of the shapes so that we could compare their recognizability. Using similar techniques, we are also studying other design elements in the road-sign context and will then move toward testing complex signs. We will also move out of the laboratory and use similar techniques to test under actual conditions of driving on the road.

Another way in which we have broadened the range of applications of signal-detectability theory is to consider the deferred-decision problem. We recognize that people do not always make decisions on a one-shot basis, as we so often call upon them to do in the laboratory. Instead, they may make repeated, independent observations of the same circumstance. Over the course of these repeated observations, there is an accrual of information such that the final decision is of higher quality than a decision could have been on the basis of any single observation. In addition, we recognize that in practical situations there may be some uncertainty as to which of a number of stimulus alterna-

tives is to be considered at any particular time.

Figure 4 represents data from an experiment dealing with these questions. The vertical axis, labeled d', represents the detectability of the signal, or, as we might say, the quality of the decision. The horizontal axis represents successive stages of observation on the part of the observers.
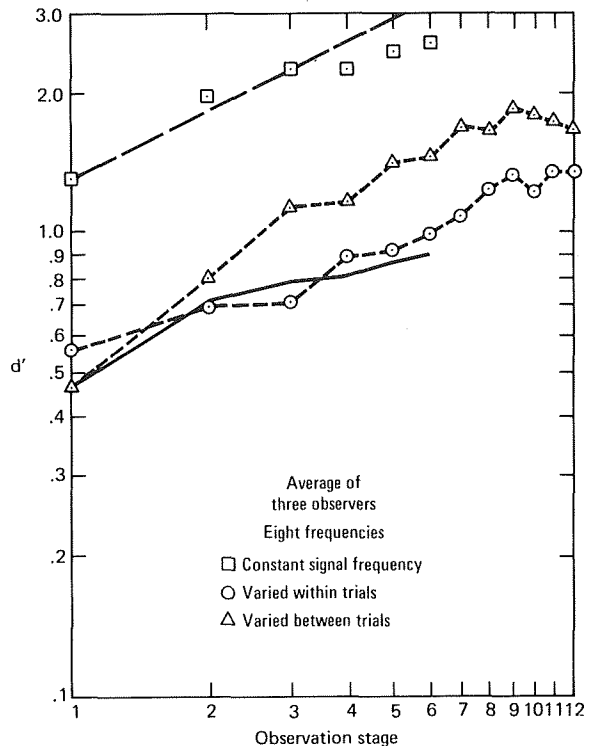

FIGURE 4.—Detectability of the signal d' as a function of observation stage for three experimental conditions. (The solid curve is a theoretical prediction.)

The actual experiment was one in acoustic detection, where observers had to detect the presence of a brief pulsed sinusoid in a background of gaussian noise. Three degrees of uncertainty were used in this experiment, the uncertainty being about what the frequency of the signal was.

In the first experimental condition, a control or baseline condition, there is no uncertainty about signal frequency. That is, the signal frequency remains fixed over the multiple, successive observations of a trial, and

over a group of trials. This is a case which we might call "signal specified exactly."

The second condition presents uncertainty, but nominally provides for no opportunity to reduce that uncertainty, nor does it permit adaptive adjustment. In this case, the signal frequency varies randomly from one observation to the next within a given trial.

The third condition presents uncertainty that could be overcome with time and is the condition of primary interest. Here, the signal frequency varies randomly between trials, but is fixed over successive observations of a given trial.

For every condition, a signal is present, or noise alone is present, in all observation intervals of a given trial. As the trial proceeds, the observer is required to respond after each observation with his best judgment about signal presence and to indicate numerically his confidence. The actual signal parameters are of little importance in this context. Instead let me discuss the data of the experiment in rather more general terms.

Consider, first, the data represented by the squares in figure 4. This is the case where the signal is specified exactly; that is, there is no uncertainty. As you can see, the quality of the decision improves with successive observations, although, in fact, the marginal improvement per unit observation seems to decrease. The dashed line is a theoretical prediction that says that the quality of the decision should improve as the square root of the number of observations. This prediction assumes that each observation is independent and as useful as any other observation. As we have noted, this seems to hold for the first few observations, but we then reach a point of diminishing returns.

Consider next the data represented by the circles. These represent the case of maximal uncertainty in the sense that successive observation cannot serve to reduce the uncertainty. The first thing to note is that there is a considerable decrement in performance as a function of the uncertainty. Note the additional number of observations that need to be taken in this condition to bring the decision quality to what it would have been

on the first observation in the case where there was no uncertainty. This turns out to be about 9 or 10 observations. The solid line in figure 4 is a fit to the circles predicted by a model, which says that the observer can listen for each of the alternatives simultaneously, but that for each alternative considered, a little more noise is introduced into the system thereby degrading the quality of the decision. The prediction is made from the no-uncertainty data represented by squares, and taking into account the number of stimulus alternatives—in this case, eight different frequencies.

The prediction seems to provide a reasonable fit for the data, and it seems tenable that an observer can pay attention to all eight frequency alternatives at one time. Later on we shall have occasion to see a case in which a similar hypothesis seems not to hold.

Now let us consider the condition of uncertainty that could be overcome by successive observations. These data are shown as triangles in figure 4. Again we note quite an initial decrement in performance in this case as compared with the condition of no uncertainty whatsoever. However, if we compare this case to the condition of maximal uncertainty, we find that the first few observations are extremely helpful in reducing the uncertainty, and that the marginal improvement in performance per unit observation is higher. If we use the same measure of decrement from the perfectly certain condition, namely the number of additional observations necessary to overcome that uncertainty, we find that in this case only four or five observations are needed. This is in contrast to the 9 or 10 needed in the maximal uncertainty condition.

So far we have covered only one aspect of the deferred-decision problem. We have left unanswered the question of who should be responsible for putting off the decision. We have dealt with this problem in the laboratory, however. Let me describe the two alternative procedures for postponing decisions that we have considered.

The first is what we might call a directive from above. In this case, the number of

observations to be taken is predetermined, and the observer simply looks that many times.

In the second case, we rely on individual initiative. That is, the observer keeps requesting additional looks until he is satisfied that an adequate decision can be made. Each of these alternatives has certain advantages of its own.

The first may be much more simply implemented. The second, as we shall see, may lead to a certain economy in the number of observations required to make an adequate decision.

Figure 5 illustrates a comparison of the two approaches. In each block, the square represents performance achieved when the initiative is left to the individual. The solid dots represent performance where the number of observations to be taken is predetermined.
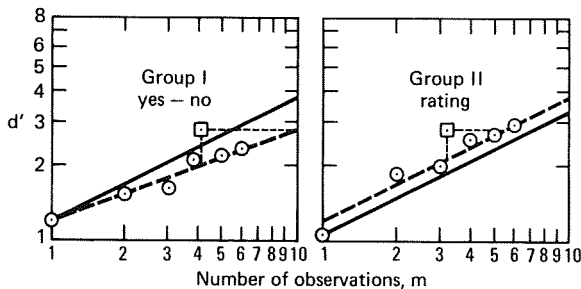

FIGURE 5.—Value of d' as a function of the number of observations for groups I and II. The figure shows average results. The circles represent data. The solid lines represent the prediction of $\sqrt{m}$ improvement in d'. The points plotted as squares were obtained with the deferred-decision procedure.

Taking our standard measure of efficiency —the number of observations that it takes to achieve a given quality of decision—we find that when the initiative is left to the observer, he requires only half the number of observations.

Thus, in practice, other things being equal, we would recommend that considerable efficiency can be gained by having the observer decide when he has enough information to make the decision. Keep in mind, however, that other things are not always equal. To provide the two-way communication necessary for an observer to request additional observations may be prohibitively expensive.

I should also explain the meaning of the two groups that are shown separately. The difference was in how the observers were trained to make decisions. On the left, the observers were trained to make decisions in a relatively coarse manner. That is, the answers that were solicited in training were purely binary responses. The observers comprising the group on the right, on the other hand, were trained to give much more finely graded responses.

As you can see from the two dashed lines fitted to the points, the accrual of information over successive observations seems much more efficient when the observers have been trained to give such finely graded responses.

We have also recognized that not only is the quality of decisions important, but also the speed with which observers can arrive at such decisions. Moreover, this aspect of performance becomes increasingly important with technological advances. Let me illustrate this point.

Advances in machine performance are almost always accompanied by increased demand on the human controller, as, for example, in the latest generation of high-performance aircraft. When demands upon the pilot become critical, a factor of prime importance is the speed with which an appropriate response can be made. In most cases, the signal of impending danger to which a response must be made is a visual one. In some circumstances a warning signal can be arranged to precede the danger signal, but, because of the inherent uncertainty in such situations, the warning signal is neither a completely accurate predictor nor can it precede the danger signal by a precisely defined time interval.

We have recently completed the first of several studies designed to shed light on what improvements in response time might be expected when such a warning signal is available. Because of the prevalence of visual

signals as warning indicators in practical situations, we have used visual indicators in the study. The danger signal was a red light, and the warning signal a yellow light. Both signals were of moderate intensity. The response used was the depression of a foot pedal.

To simulate the practical use of such a warning indicator, we used a warning light that had the two properties previously discussed. First, it was not always followed by the red light to which a response was to be made. Instead, the probability with which a red light followed the yellow light was a parameter of the experiment. The second property of the warning light was that the interval by which it might precede the danger signal was a variable one. This interval was also a parameter of the experiment.

For purposes of comparison, we did not use a yellow light for half the sessions. In those sessions that did not use the yellow light, the frequency of the red light coming on was varied over the probabilities, 0.2, 0.5, and 0.8.

The duration of the yellow light took on eight different values with equal likelihood. In the comparison session, the duration of yellow on-time is also treated as a variable, even though the subjects could not see the yellow light. It should be noted that one of the intervals used was 0 second. In this case, there was no warning signal. As we shall see, this 0-second case is not identical with performance in those sessions where no warning light at all was used. In the 0-second case, the subjects were expecting the warning light.

The first finding is illustrated in figure 6. All data are the averages of the individual subject medians, and in each case the graphs are representative of individual subject performance. In figure 6, the reaction time to the red light is plotted against the relative frequency of the red light. In the case where the yellow light was inoperative, given by the dashed line, the reaction time is seen to decrease slightly, simply as a function of the increased density per unit time of the occurrence of the red light.
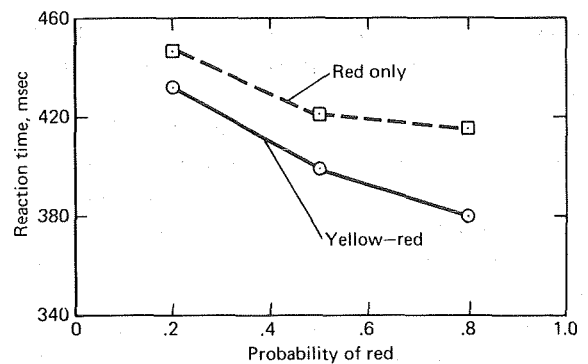

FIGURE 6.—Reaction time as a function of the probability of occurrence of a danger signal (following the warning signal—solid line).

The case where the yellow light was operative, indicated by the solid line, shows a steeper decrease, reflecting not only a decrease in reaction time due to the increased density of red-light occurrences, but also reflecting a decrease in reaction time due to increased probability that the yellow light would be followed by a red light. Figure 6 also indicates the decrease in reaction time that can be achieved by preceding the visual signal with a warning signal.

In figure 6, the data arising from the condition of zero on-time for the yellow light are purposely omitted. The reason for this omission is illustrated in figure 7, which shows the reaction time as a function of the duration of the yellow light. The probability
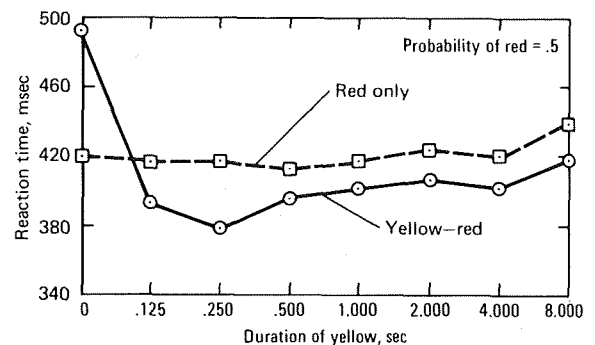

FIGURE 7.—Reaction time to a danger signal as a function of the interval between the warning and the danger signals.

of occurrence of the red light for these data is 0.5, and they are representative of the data at the other probabilities.

As expected, the reaction times are constant when the yellow light was inoperative. In the case where the yellow light was operating, the reaction times are everywhere faster except where the yellow light did not precede the red light—that is, at the zero interval. In fact, the reaction times at the zero interval are quite a bit longer when the yellow light is operative in this case as compared with the case when no yellow light is used.

That is, of course, a reflection of the usefulness of the warning light and the subjects' reliance upon it. It does, however, point up a hazard that must be faced when an early-warning system is implemented. In a practical application, the zero interval would represent performance in the event of a failure of the warning indicator. As a consequence, we emphasize the need for high reliability, perhaps achieved through redundant encoding for such an indicator.

To summarize our findings here, the value of such an early-warning indicator would be greatest when it is followed quickly, and with high probability, by the signal to be responded to, and we should reemphasize the necessity for high reliability of the warning indicator.

We are also mindful of the fact that, in many practical situations, the visual and the auditory systems of humans may be overloaded. For this reason, we have begun to explore a range of vibratory stimuli. Within the aeronautical context, it is an interesting sidelight that, having designed out the seat-of-the-pants feel from aircraft, we may end up designing back in pertinent vibratory stimuli.

There are other reasons, too, aside from the need for a new information channel, that compel us to consider vibratory stimuli. There is, of course, the desire to investigate a new sensory modality per se, to find out how it works and how it compares with other modalities.

There is also the desire to demonstrate the generality of our theoretical approach. If we can show theory-bound similarities in the quantification of one sensory modality to another, then, even though in quality they may be quite disparate, we may be able to treat them as interchangeable subsystem components.

We undertook an initial investigation to see if vibratory stimulation to the fingertips would summate; that is, if detection is better when the signal is applied to two fingers than when the signal is applied to only one. Figure 8 shows that the proportion of correct responses (in a two-interval forced-choice test) was no greater if two fingers were stimulated than if only one was.

A model consistent with this behavior is that the observer can attend to inputs to a single channel only—in this case, a single finger—at any given instant. This model is in contrast to the multiband model that we applied to the problem of deferred decision with uncertainty.

I should mention that this experiment did not involve uncertainty. The observer knew which of two, or both, fingers would be stimulated.

Figure 9 shows what happens when there is, in fact, uncertainty about which finger or fingers would be stimulated. Here we see that performance for two fingers is superior. While this might be interpreted as summation, it is consistent with the single-channel model. The interpretation here is simply that when a single finger was stimulated, the observer might well have been paying attention to the wrong finger. If both fingers were stimulated, then he was certainly paying attention to a relevant finger. Thus performance would be superior when the two fingers were stimulated.

By no means is all of our work of an applied nature, nor does it permit direct and immediate application. Much of our work is oriented toward future applications—perhaps the distant future. For a fraction of our work we see no applications; however, we trust that others will.
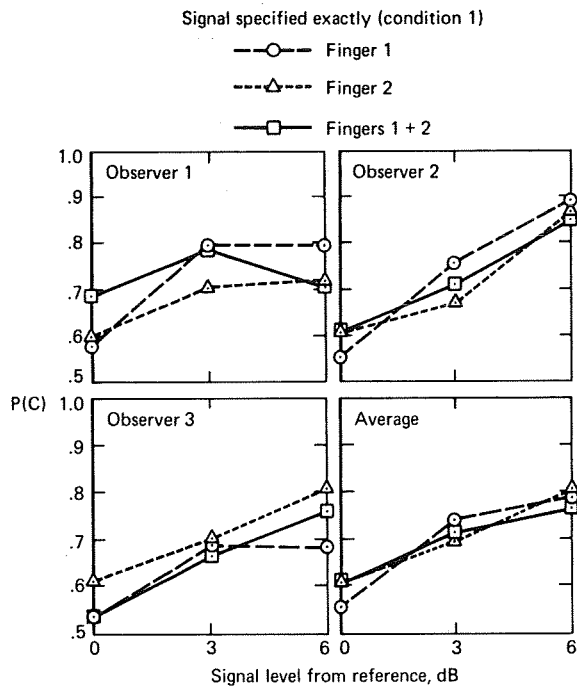
FIGURE 8.—P(C) versus signal level for three observers, and the average for finger 1, finger 2, and fingers 1+2, with the signal specified exactly.
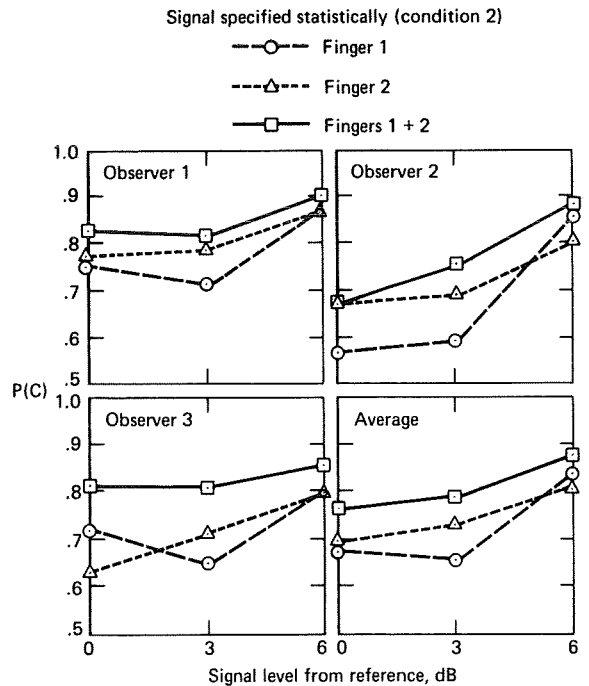
FIGURE 9.—P(C) versus signal level for three observers, and the average for finger 1, finger 2, and fingers 1+2, with the signal specified statistically.

DISCUSSION

LLOYD A. JEFFRESS: Did the observer know that both fingers would be stimulated in both cases?

JOSEPH MARKOWITZ: In figure 8, the observer knew what finger or combination of fingers would be stimulated, so there was perfect certainty. In figure 9 he knew that either one or both would be stimulated, so he knew that was a viable option.

ALFRED B. KRISTOFFERSON: Were these brief signals?

MARKOWITZ: No. We started with brief signals of 100-millisecond duration, then switched to 500-millisecond duration. So these are 500-millisecond signals in both figures 8 and 9, not what I call brief.

JEROME I. ELKIND: What kind of vibration?

MARKOWITZ: 250 Hz, more or less.

Visual and Auditory Signal Recognition[1]

LLOYD A. JEFFRESS

Defense Research Laboratory
The University of Texas

VISUAL RESEARCH

Effects of Chromatic Adaptation on Color Naming

The four color names—blue, green, yellow, and red—were employed singly or in pairs by the subjects in identifying the color presented to them after a period of chromatic adaptation. The responses were scaled as follows: blue was graded 3; blue-green was scaled 2 for blue and 1 for green; green-blue was scaled 2 for green and 1 for blue; and so forth. There was a possible total of 72 points for each test wavelength. The graphs show the percentage of total points assigned to each wavelength indicated along the abscissa. Figure 1(*a*) is the results for three subjects after 5 minutes of initial neutral adaptation at 195 ft-L. The test stimuli were of the same luminance and were presented for 300 milliseconds at 18-second intervals, alternating with the adaptation light that was on between test trials. The stimuli were presented in Maxwellian view subtending 40°.

Figure 1(*b*) shows the results after adaptation with a W 92 filter (646 nanometers). The red has almost disappeared, the yellow is

shifted well toward the red, the green has been extended over a wide range of wavelengths, and the blue is virtually unaffected.

Figure 1(*c*) shows the effect after adaptation with a W 98 filter (452 nanometers). Here the blue has been greatly restricted and moved to the left. The red has shifted to the left and even appears in the blue region as red-blue or blue-red. The extent of the green has been greatly reduced and shifted to the left, and the area of yellow has been increased and shifted to the left.

Figure 1(*d*) shows the effect of adaptation with a W 74 filter (538 nanometers). Here the red and blue have been expanded toward the middle, and the green and yellow responses completely suppressed for one subject and greatly restricted for the other.

The results, in addition to indicating the effect on hue of prior adaptation, illustrate the effectiveness of color naming as a quantitative experimental research procedure. Split-half reliability correlations for the data were mostly in the high 90's, and the method is much less time consuming than matching procedures.

One practical suggestion from the results concerns the use of colored light in the illumination of sonar and radar spaces. The red commonly employed on board ship to preserve dark adaptation is about the most inappropriate lighting when the color to be detected is the greenish-yellow of scope phosphors. Either neutral light or blue would be much better.
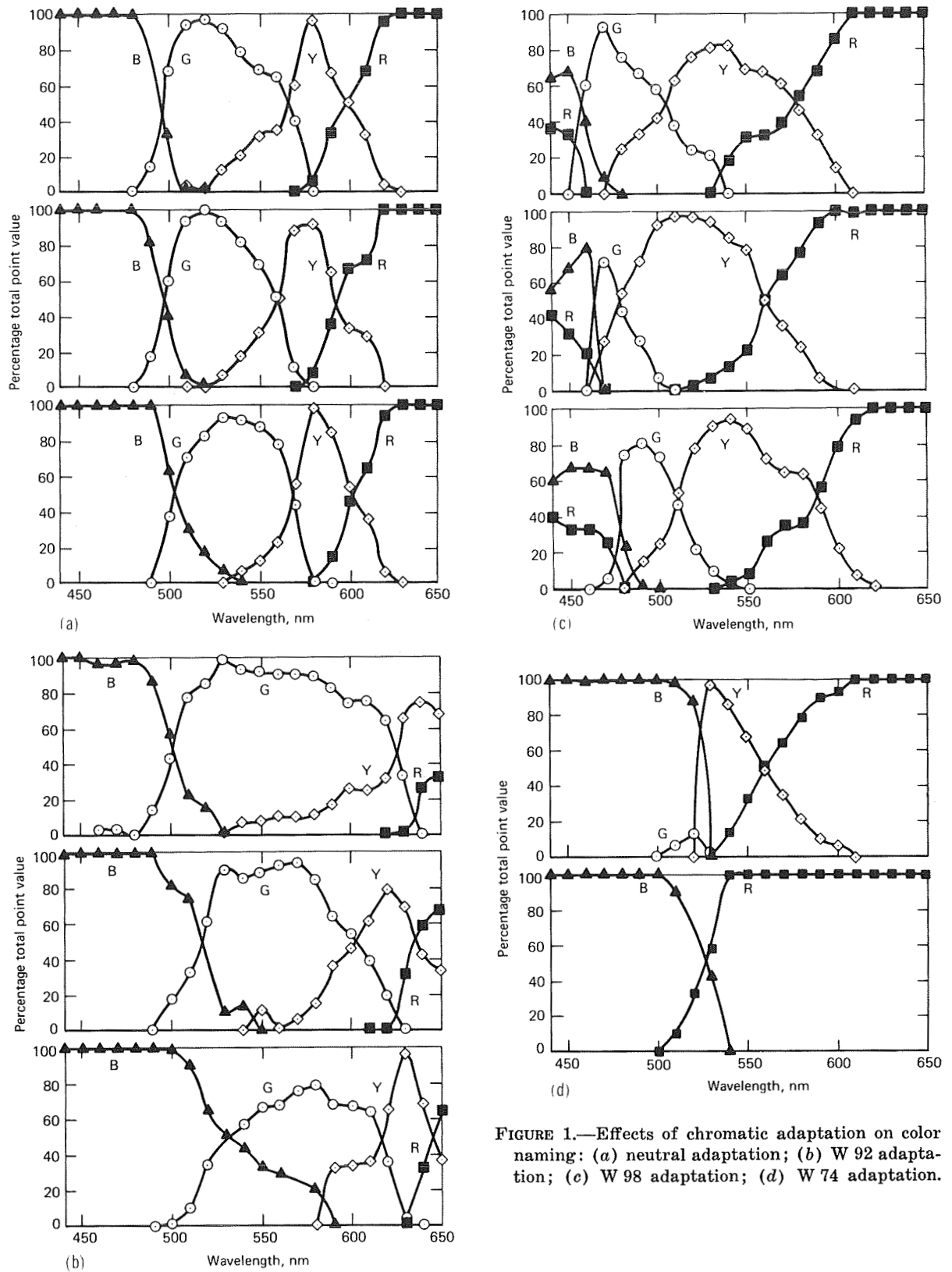
FIGURE 1.—Effects of chromatic adaptation on color naming: (*a*) neutral adaptation; (*b*) W 92 adaptation; (*c*) W 98 adaptation; (*d*) W 74 adaptation.

Saturation Estimates and Chromatic Adaptation

In addition to assigning color names reliably, subjects prove to be able to estimate the saturation of colors presented after various types of adaptation. The subjects were instructed to assign numbers ranging from 0 to 10 to the saturation of test colors presented after adaptation. Figure 2(a) shows data for three subjects after neutral adaptation for 5 minutes (Maxwellian view, 40°, same procedure as in the previous experiment). The bars indicate one standard deviation above and below the mean estimate.

Figure 2(b) shows the effect of adaptation to a long wavelength (W 92, 636 nanometers). Note that saturation estimates for the long wavelengths have been greatly reduced. Figure 2(c) shows the effect of adaptation to a short wavelength (W 98, 452 nanometers). Here there is a marked reduction in saturation for wavelengths from medium to short, with increased saturation for wavelengths at the other end of the scale. Figure 2(d) shows the effect of adaptation for green (W 74, 532 nanometers). There is depression of the estimates in the middle, with a considerable increase in the estimates for the long wavelengths, and, for one subject, for the short as well. The relatively small spread for any particular wavelength indicates that the judgments are being made with consistency.

Effects of Adaptation on Visual Detection

It is frequently suggested in recent literature (but originally by Barlow in 1964) that the effect of exposing the retina to light is to make the retina's behavior in subsequent darkness noisy. This implies that the detection of a weak visual signal in the dark following a brief flash of adaptation light is essentially the detection of a signal in noise. For this, and a number of other reasons, the present experiment was undertaken as a detection task. A rating-scale procedure was employed to permit the construction of receiver operating characteristic (ROC) curves from the subjects' responses.

The adaptation light and the signal were presented in Maxwellian view. The former

(a)
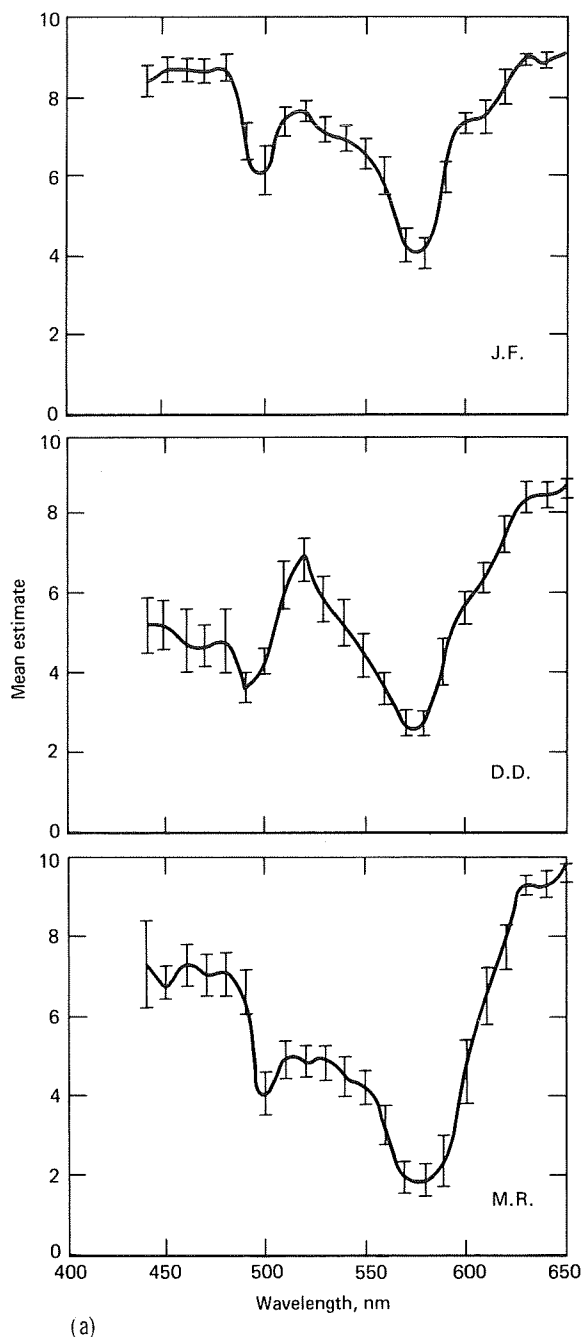
FIGURE 2(a).—Saturation estimates and chromatic adaptation: neutral adaptation.

subtended an angle of 25°; the latter subtended an angle of 5° in the center of the adaptation field. The exposure of the adaptation light was controlled by a mechanical shutter that allowed the light to be presented
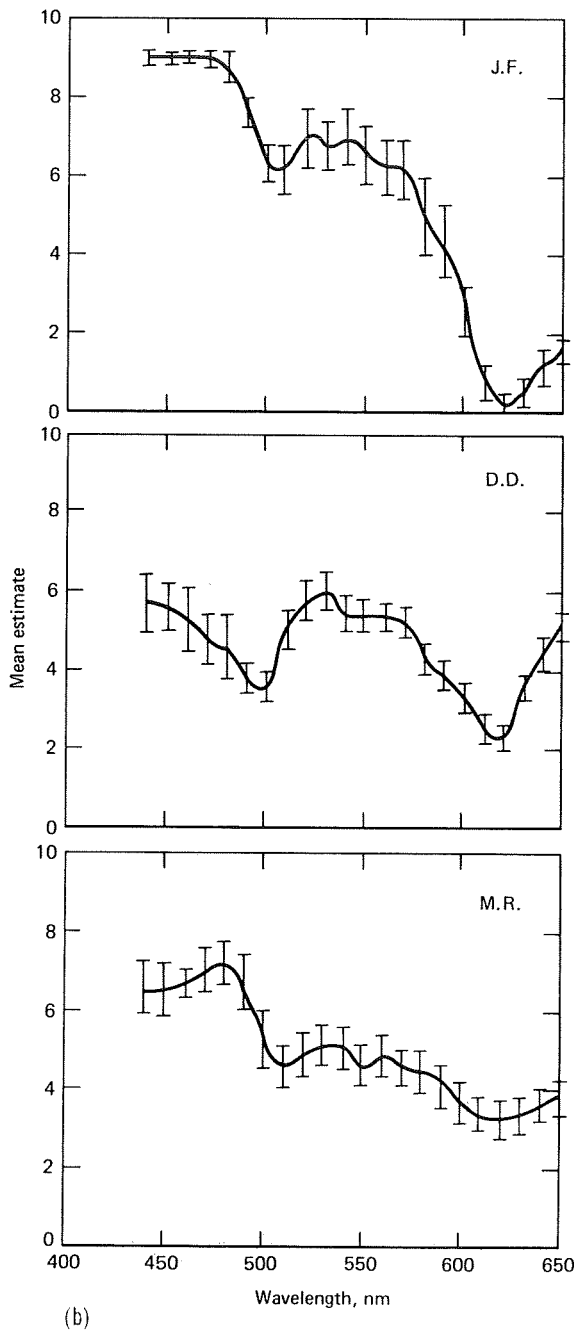
FIGURE 2(b).—Saturation estimates and chromatic adaptation: W 92 adaptation.

FIGURE 2(c).—Saturation estimates and chromatic adaptation: W 98 adaptation.

for 200 msec. The retinal illuminance provided by the adaptation light was 6.74 log trolands. The test light (signal) was a glow-modulator tube, which was flashed electronically for 20 msec at a constant illuminance level (constant spectrum) and attenuated by a series of neutral filters ranging from 2.04 to 0.54 log trolands of retinal illuminance in 0.1 log unit increments. A bite board and a dim red "grain-of-wheat" fixation light
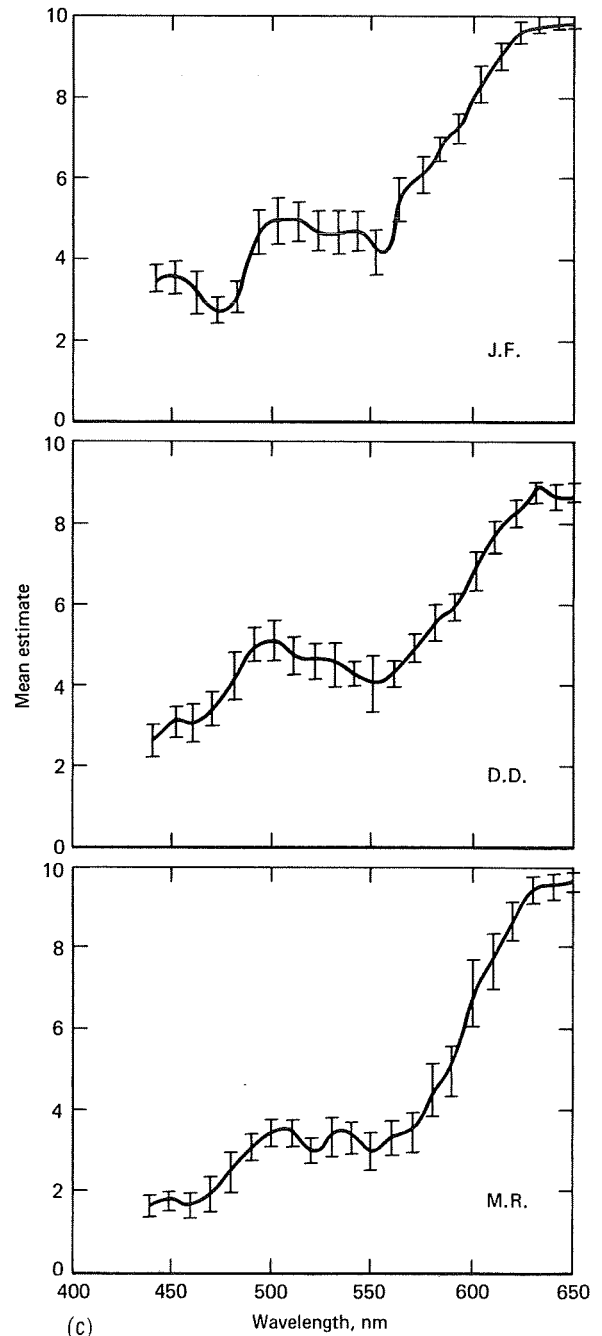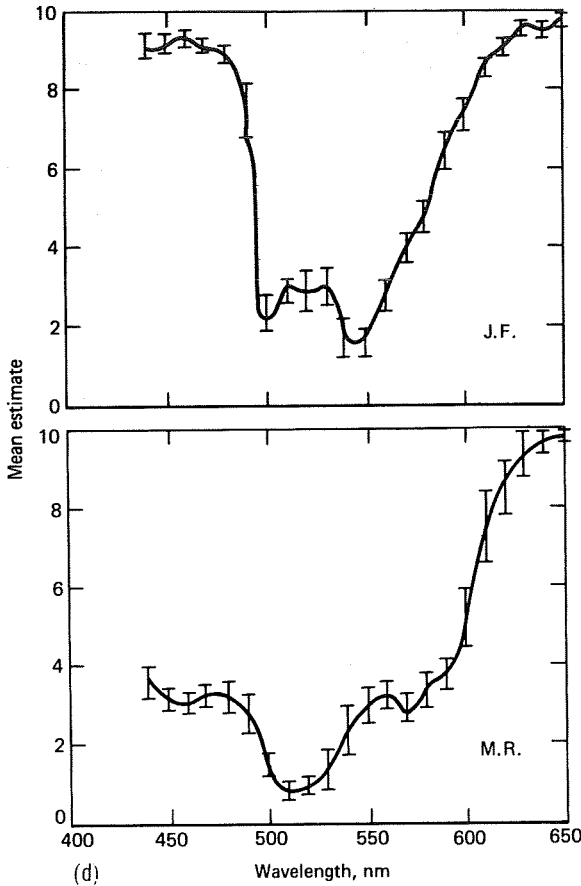
FIGURE 2(d).—Saturation estimates and chromatic adaptation: W 74 adaptation.

maintained the desired orientation of the eye.

After spending a minimum of 10 minutes in a dimly lighted room, the subject entered the dark test booth. After 2 minutes of dark adaptation, the fixation light was turned on and he signaled by means of a pushbutton that he was ready. The adaptation light was then turned on for 200 msec, and every 6 seconds following the termination of the adaptation light, a 1-second warning tone was presented. At the termination of the tone, the test signal was either turned on or was not (with an a priori probability of 0.5). The subject responded with an appropriate pushbutton to indicate his assurance that a signal had or had not been presented. A response of 1 represented virtual certainty that there had been no signal; 10 represented

virtual certainty that there had been a signal. Forty such 6-second periods following the adaptation flash constituted 1 run, and 5 to 10 such runs, separated by 2 minutes of dark adaptation, constituted an experimental session. The luminance of the test signal was varied during the run and from one run to another, according to a planned-haphazard program, so that any of 6 or 7 illuminances might occur during any trial in the run of 40. The values were chosen so that the percentage of correct responses at any period following the adaptation flash fell within a reasonable range for getting ROC curves. Some 400 ROC curves were obtained during the course of the study and the values of P_c were determined from their area as measured by a planimeter. Figure 3 shows

FIGURE 3.—ROC curves for 0.54 log troland "signal."
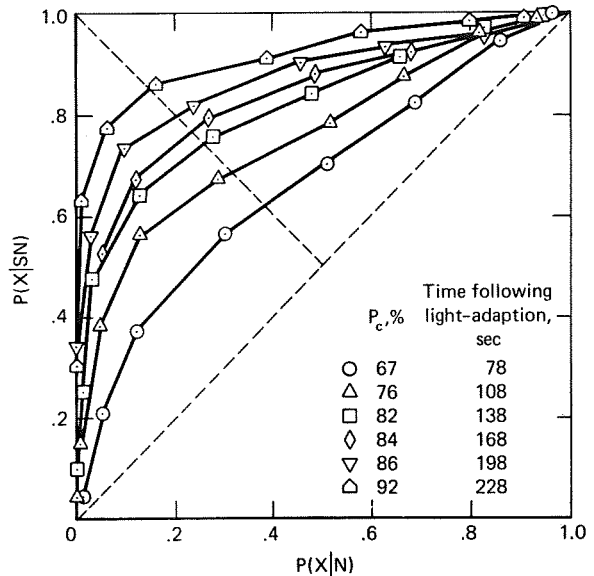
a family of ROC curves for the average of the three subjects. These data were taken at a test-light illuminance of 0.54 log troland. The parameter of the family of curves is the time following the adaptation flash at which the data were taken.

Figure 4 shows the course of dark adaptation. The abscissa is time after the adaptation flash, and the ordinate is the illumi-
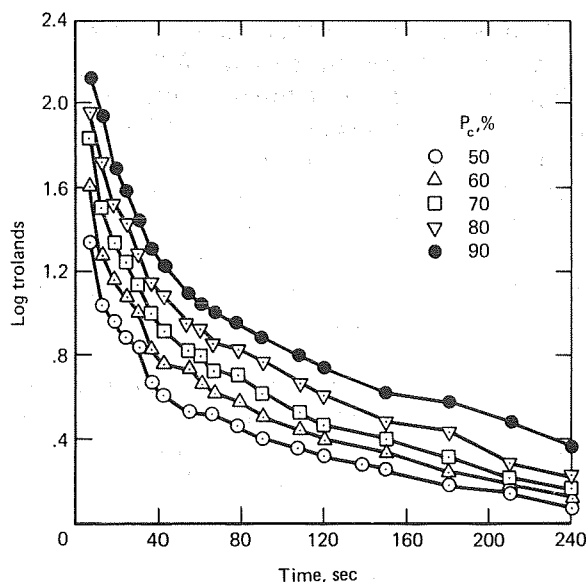
FIGURE 4.—Course of dark adaptation for one subject.

nance needed to reach the percentage correct, indicated as the parameter of the family of curves. Figure 4 is for a single subject. The others showed similar results. The subjects were asked to report the moment when the positive afterimage of the adaptation light disappeared. This occurred about 2 minutes after the exposure, but was not accompanied by any discontinuity in the recovery curve. The retinal noise apparently does not depend on the presence of the positive afterimage.

This study is forming the basis for a doctor's dissertation by Mr. Heinz Gaylord.

The Bezold-Brücke Hue Shift

Although the Bezold-Brücke shift is important to theories of vision, much of our knowledge of the phenomenon is based on a study by Purdy, made over 30 years ago on a single subject. It therefore seemed appropriate to examine the effect with a substantial population. In the present study, 72 subjects (33 females and 39 males) were employed. The color-naming procedure described earlier was used. The apparatus was a two-beam device, with one beam providing the low-level adaptation light and the other, the test wavelength. Measurements were made in the range from 470 to 630 nanome-

ters, using a grating monochromator adjusted to yield a passband of 15 nanometers. Two luminance levels, 320 and 3200 trolands, were used with central fixation in Maxwellian view subtending an angle of 3°. Test stimuli were presented for 300 msec once every 18 seconds in random presentation order. The adaptation light was viewed during the times intervening between test stimuli. Every subject was given several practice stimuli before data collection was undertaken. Each subject served for 1 hour and received as many stimuli as could be programed in that time.

The color-naming values were converted into nanometers of shift and are presented in figure 5. The plotted points are the means for the sample of observers tested, and the bars represented two standard errors of the mean for the point. In general, the shifts shown here are smaller than those reported by Purdy, but the so-called invariant points occur at 584, 502, and 474 nanometers for the mean of the present sample—about the same locations as found in earlier studies.
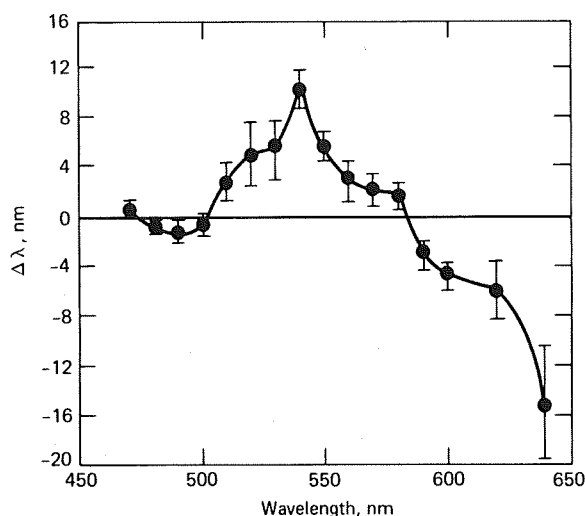

FIGURE 5.—Bezold-Brücke hue shift.

The variability shown in figure 5 probably is largely the result of individual differences, because an earlier study shows that the method can be highly reliable.

It was also possible from the data to

determine the spectral location of unique yellow and unique green at the two luminances employed. The location of the unique yellow did not change systematically. The mean was 580.55 nanometers under the low-luminance condition and 580.50 nanometers for the high. However, the location of the unique green showed an interesting effect. Table I shows the finding.

TABLE I.—*Unique Green Loci (in nanometers) for 2 Classes of Observers at 2 Luminance Levels. Results of Statistical Evaluations of Row and Column Differences Are Indicated.*

Group	N	Low luminance	High luminance	Difference
I	19	513.1	509.9	$p<0.05$
II	11	525.5	511.0	$p<0.01$
		$p<0.01$	$p>0.05$	

The subjects appear to fall into two groups in their location of unique green at a low luminance level. The difference disappears at the high luminance level. Both groups show a shift in the location of the unique green at the low level, but the shift for group II is much greater than for group I.

In addition to the psychophysical work just reported, Jacobs and his students are conducting behavioral and neurophysiological studies on animals, with additional support from the National Science Foundation. This work involves both color and spatial sensitivity of single units of the lateral geniculate. The findings of the work on color sensitivity are shown to correlate highly with the behavior of the animals in discrimination tasks.

AUDITORY RESEARCH

Beginning in May 1964, the National Aeronautics and Space Administration provided support for work in audition already receiving support from the U.S. Navy Bureau of Ships. The addition allowed us to increase our efforts in this field and to provide assistance for more graduate students, both experimenters working on dissertation problems, and subjects who receive hourly pay for their services. The addition also made it possible for us to construct considerably more flexible programing and recording equipment. At the present time, most of our data are recorded on punched cards that are then analyzed by means of the laboratory's CDC 3200 computer. Four subjects at a time can be run in psychophysical studies. Several signal levels can be employed in a single session, and single-interval, two-interval, forced-choice, or rating-scale responses can be recorded. Where desired, examinations of serial effects and multiple-observer responses can be made.

Most of our recent work in audition has been concerned with detecting a signal in noise, although earlier a considerable amount of research was devoted to various problems in the localization of sound. The masking studies have fallen into two main categories: those concerned with the detection performance of the single ear, and those concerned with the binaural release from masking, which can occur when stimuli to the two ears are not identical.

BINAURAL STUDIES

Time and Intensity Differences and Lateralization

This was a study conducted by a Summer, Science-Participation, high-school student (Brant T. Mittler) under National Science Foundation support, and supervised by Dr. Charles S. Watson. The student and his subjects were 17-year-olds. The subject's task consisted of drawing lines across a sketch of the head to indicate the range of movement of a commutated sound. The sound, a 500-Hz tone, was presented via earphones with either a level difference or a phase difference between the inputs to the phones. The inputs were commutated at half-second intervals and produced a distinct impression of movement within the head. The locations of the ends of the lines represented where the subject thought movement began and ended. The data sheet was located behind a slit in a sketch of a face and moved between trials so that each judgment could be made without

reference to previous ones. Figure 6 shows the mean length of line associated with the intensity difference or the time difference shown on the abscissa. The "trading ratio" obtained in this way agrees with others in the literature, about 60 μsec/dB.
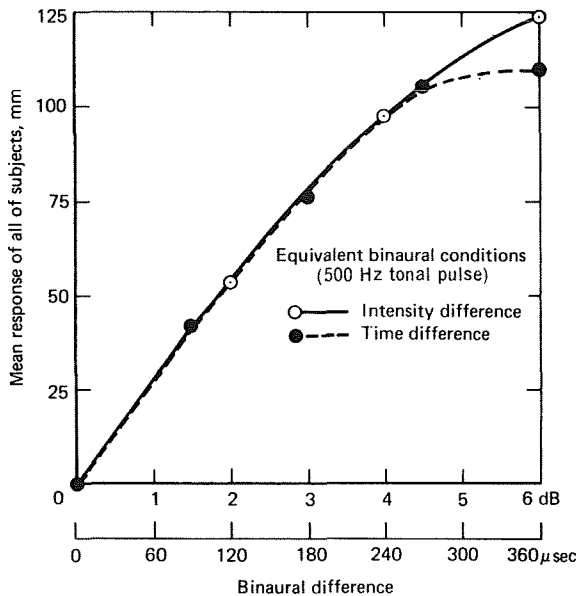

FIGURE 6.—Length of lateral movement of a sound.

Masking-Level Differences for Tone and Narrowband Noise

A 50-Hz wideband of noise centered at 500 Hz, and a 500-Hz tone were employed as signals in a binaural masking experiment. Hirsh and Webster had reported much larger masking-level difference for a noise signal than for tone. The present experiment was undertaken, in part, to check their findings for which no theoretical explanation was apparent. The major results are presented in figure 7. There is no significant difference between the masking-level differences for noise and tone, and it makes little difference whether the noise is shifted in time, by a delay line, or in phase by a phase shifter. A second experiment was an attempt to explain the findings of Hirsh and Webster and revealed that they had employed the same noise generator for their masker as for their signal. When these conditions were repli-
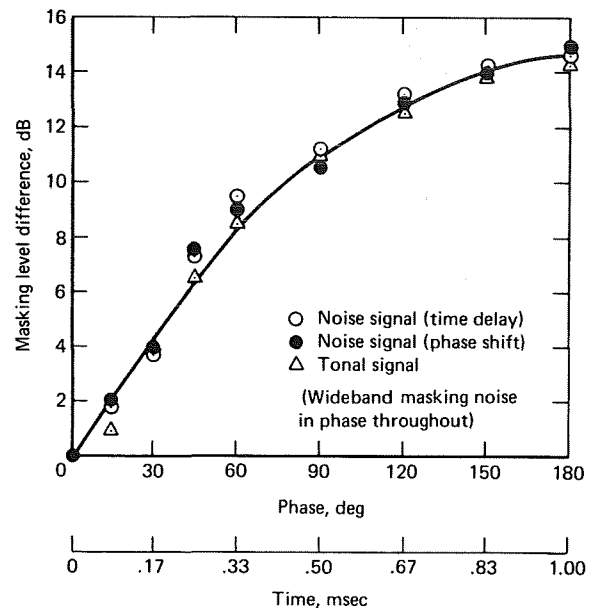
FIGURE 7.—Masking-level differences for tone and narrowband noise.

cated, our findings agreed with theirs. A large masking-level difference (18 dB) was obtained when the masker and signal were in phase opposition. It occurred, however, because of the considerable increase in the signal needed in the NO SO reference condition, not because of any great release from masking under the antiphasic condition.

Binaural Detection as a Function of the Bandwidth of the Masking Noise

Earlier work has suggested that the bandwidth involved in binaural detection is somewhat wider than that for monaural detection. The present experiment was undertaken to study this possibility. Equivalent rectangular bandwidths of 2900, 508, 422, 303, 185, 160, 130, 109, 50, 22, and 12.6 Hz were employed for the masker. The signal was a 500-Hz tone of 150 msec duration and a rise-fall time of 25 msec. Three levels of noise were employed: 50, 45, and 30 dB spectral level. A low-level, wideband background noise was used to mask the second harmonic of the signal, which was about 60 dB below the fundamental. The stimuli were presented either with both noise and signal in phase at

the two ears, NO SO, or with the signal reversed in interaural phase, NO $S\pi$. Figure 8 (a) shows the results for one subject for the diotic condition NO SO. It will be seen that the effect of band narrowing is not very

(a)

(b)
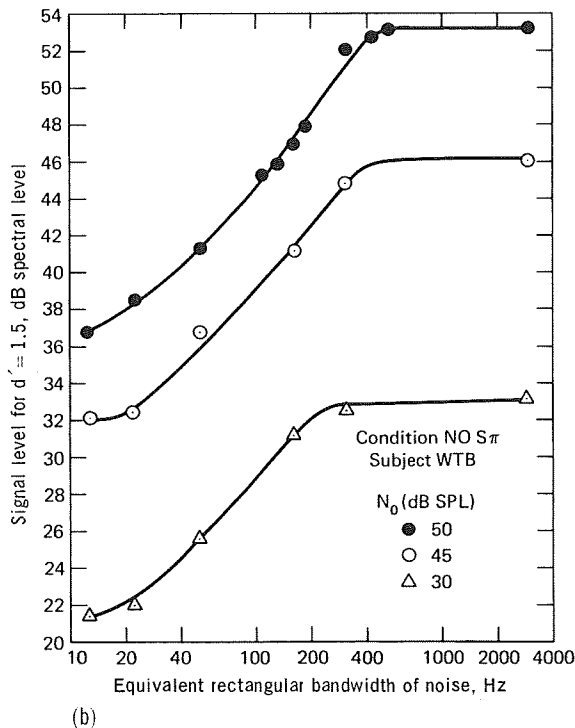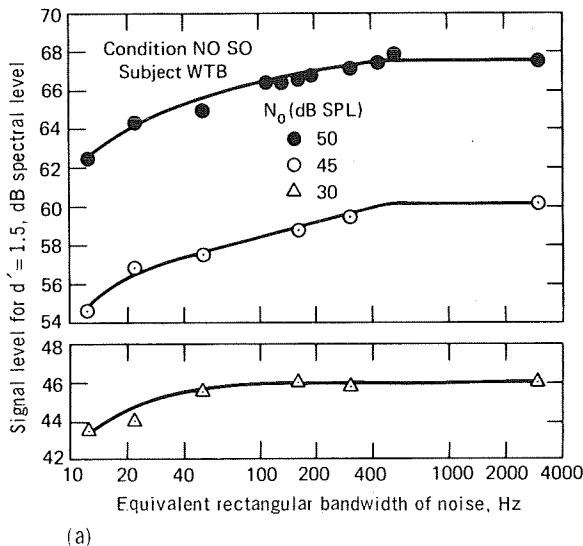
FIGURE 8.—Effect of bandwidth on homophasic and antiphasic detections. (a) Homophasic, (b) antiphasic.

significant, until a bandwidth of about 50 Hz is reached; whereas, figure 8 (b) shows very substantial improvement beginning at bandwidths as wide as 200 Hz. The results strongly suggest that a much wider range of frequencies is involved in the detection of a 500-Hz tone under the NO $S\pi$ condition than is involved in monaural or NO SO detection. This is probably not surprising, because we are presumably concerned with a population of auditory nerve cells that are different in binaural phenomena from these for monaural detection. Neural "funneling," as Békésy calls it, probably occurs in narrowing the bandwidth for monaural detection, whereas probably only the filtering provided by the mechanical action of the basilar membrane determined the bandwidth for binaural detection.

Binaural Electrical Models and Detection

In an attempt to replicate human performance, we have tested several electrical models of the binaural detection mechanism in psychophysical experiments. Two such models have been run as subjects along with three human observers in a 2AFC experiment. The first model converts the interaural time difference produced when a signal is added antiphasically to an in-phase noise into a voltage that is averaged and sampled at the end of the observation interval. To avoid perfect performance when the noise is in phase at the two ears and the signal reversed in phase at one ear, a small amount of uncorrelated noise is introduced into one channel of the model. This simulates the "noisiness" of the subject's transduction of waveform into nerve impulses. The model yields psychometric functions that fit human functions either at high signal levels or at low, depending on the noise correlation used. It has not been possible with this device to fit human performance over the whole range of the psychometric function. This fact may be the result of a major inadequacy of the model. It takes into account the time differences based on axis crossings, but fails to make use of level differences that result from adding the signal to the noise. A second

model, based on Durlach's equalization-cancellation model, has so far failed to perform as well as the one just described. A third model employing the cross correlation between the two earphone channels will be similarly employed if some of its present weaknesses can be eliminated.

MONAURAL PHENOMENA

Effect of "Vigilance" in an Auditory Detection Experiment

Many attempts to improve detection by manipulating the values and costs matrix have failed to produce an appreciable improvement in detection over a block of trials. The present experiment was undertaken with the idea that enhanced vigilance can be maintained for only a short stretch of time. Accordingly, certain trials were selected as the "important" trials and their presence was signaled to the subject by a light. In the first experiment, the subject was told that these were the important trials and that they must be particularly careful to respond correctly (in a 2AFC setting). This preliminary experiment failed to reveal any improvements on the important trials. The next experiments involved various schedules of punishment for incorrect responses on the indicated trials. Punishment was a mild shock (1.6 milliamperes) applied to the ankle, and the experiments differed in the number of successive trials that were included in the critical block. Figure 9 shows
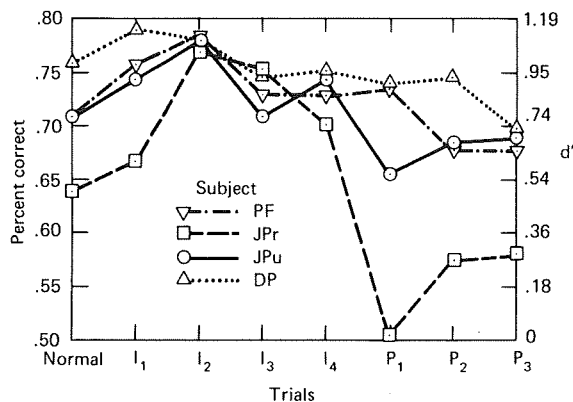
the results of an experiment where the number of important trials was four. The shock for an incorrect response could occur in any of the four trials. The results show a substantial improvement by the second trial, but a falling off after that. The postshock trials showed a considerable decrement for two of the subjects, with a gradual return to a normal level of performance. The first points on the graph are the average for the preceding 16 days of training without the interpolated trials. The findings show that improved detection can be achieved for a very short time but is not maintained. The average for the whole block of trials was the same with and without the shock.

Width and Shape of the "Critical Band" Involved in Masking

Considerable disparity exists in the literature between the estimates of the "critical band" width. We undertook the present study to obtain a better idea of both the width and the shape of the band of frequencies involved in masking a 500-Hz signal. The experiment used a set of high-pass and a set of low-pass filters in order to approach the signal frequency from one side at a time. The results, which are being prepared for publication, show that the shape of the ear's filter is distinctly asymmetrical (having much higher skirts on the low-frequency side than on the high) and that the equivalent rectangular width is of the order of 50 to 80 Hz. One important finding appears to be that subjects differ in their bandwidth. One subject who performed more poorly than the others began to improve at considerably wider bandwidths than the others. That is, he required less narrowing of the masking noise to show improvement than the others did. Apparently, in experiments where the task is the detection of a tonal signal, the Fletcher-type estimates of bandwidth are appropriate.

Models: Electrical and Mathematical

The mathematical theory of signal detectability (TSD) is based, in the usual derivations, on sampling theory—on taking a series

FIGURE 9.—Effect of punishment on detection performance.

of 2WT samples of noise (N) or noise plus signal (SN), where T is the temporal duration of the sample. There is some confusion about the meaning of W. Some writers treat it as if it were the bandwidth of the masking components of the noise (the critical bandwidth); others treat W as if it were the bandwidth measured from zero; that is, as if it were the highest frequency present in the noise sample. In any case, N and SN are sampled in the same way and for the same duration. These assumptions immediately lead us into trouble when we attempt to apply the theory to human observers or to electrical models. The assumption that N and SN are sampled similarly means that they are sampled after being filtered; that is, that the gate follows the filter. In hearing, the filtering is presumably being done by the ear and the gating in advance of the earphone. Thus, the transient responses of the filter become involved. The mathematical theory neglects this aspect of hearing.

Also, when we consider the common experimental condition where the noise is continuous and only the signal is gated for a time T, we are forced by mathematical theory to assume that somehow the subject is able to gate the noise in the same way that the experimenter gates the signal—not a very realistic assumption. The question of when and how long to sample becomes one of major concern when dealing with a physical model of the auditory system.

The Role of Signal Duration

The classic study of the role of duration in the detection of a gated signal in a continuous noise background by Green, Birdsall, and Tanner in 1957 employed a constant-energy signal of various durations and used a four-interval forced choice procedure for determining the observers' d''s. The basic finding was that observers did best over a range of durations from about 20 to 200 msec and fell off rather sharply for durations much longer or shorter than these.

We attempted to replicate the results on signal duration by an electrical model that consisted of a narrow filter, a half-wave

rectifier, and a postdetection (envelope) filter. When the postdetection filter had the short time constant needed to obtain a close-fitting envelope, the data failed to resemble that of the experiment by Green et al. Instead of being reasonably flat across a range of durations, the data showed a decided peak at a duration that was the reciprocal of the filter bandwidth. Only when we increased the time constant of the postdetection filter to 50 or 100 msec did we succeed in replicating the psychophysical data. This time constant is of the same magnitude as that obtained by Zwislocki from a very different set of experiments.
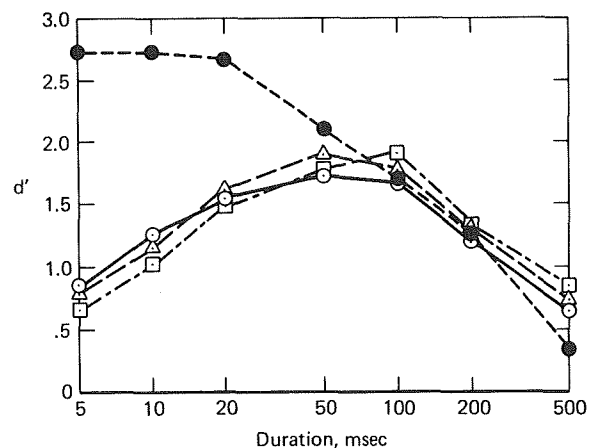
FIGURE 10.—Effect of signal duration on detection.

Figure 10 shows the results of the final series of experiments. The circles show the averages for the subjects of the experiment by Green et al., the triangles show the data obtained with the model using a half-wave rectifier, and the squares show the effect of employing a square-law (energy) detector instead of the half-wave. Closer observation of the tenets of TSD, by gating both the masking noise and the signal in the same way, obtained the solid circles. This suggests that if subjects are presented with gated noise and signal they should perform better, for a constant energy signal, when both the noise and signal are gated than when the noise is continuous and only the signal is gated.

Gated Noise and Signal

Following the lead suggested by figure 10, we undertook an experiment in which subjects were presented signals of various duration but constant energy with continuous noise and with noise gated for the same duration as the signal. Figure 11 shows the
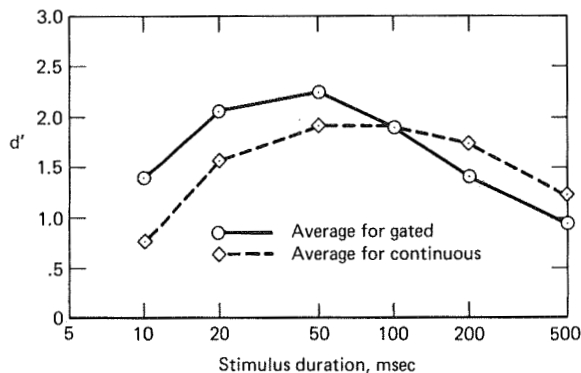
FIGURE 11.—Detection for continuous and gated noise.

averages for three subjects. Note that the subjects performed better with gated noise than with continuous noise. They did not, however, show, as the model did, continued improvement for gated noise and signals at the short durations.

John Whitmore (a postdoctoral student here) suggested that detection of a signal in a brief gated noise is a very difficult judgment, and that possibly our findings would be different with highly trained subjects. The experiment was therefore repeated using trained observers, with the result that the predicted improvement in performance as the signal was shortened was actually observed. The subjects did better at 5 msec than at 10 and better at 10 msec than at 20 or 50. The results are being prepared for publication.

Electrical Model as a Predictor of Observers' Responses

Because the electrical model appeared to simulate human performance in several important respects, it was employed as a sub-

ject along with human observers in several psychophysical experiments. In the first experiment (by Thomas L. Nichols), the model was run as a subject along with a human observer in a yes-no experiment (four subjects were tested in this way). It proved to predict the subjects' responses better than whether the signal was present or not. It also proved to be a better predictor than another electrical measure of the stimulus. This was a peak device that recorded the largest envelope peak that occurred during the 250-msec observation interval. Both noise and signal were gated for 250 msec. The two electrical measures showed a correlation of 0.5 to 0.6 for the 250-msec duration. Shorter durations increased the correlation to near unity for very short durations. The 250-msec duration was chosen to permit the two electrical measures to be reasonably independent with the possibility that they would respond to different aspects of the stimulus and predict the subjects' responses better than either measure alone. Actually, the peak detector added only about 1 percent to the predictions of the other electrical model.

A second experiment with the model was carried out—this time employing it along with three human observers in a 2AFC experiment using seven levels of signal, and some trials on which noise alone was presented in both intervals. Table II shows the results.

The first column is the signal employed, ranging from an E/N_o of 12.8 to 0—noise alone presented in both intervals. The second column is the percentage correct for the model, and the third is the average percentage correct for the three subjects. The model yields superior detection throughout the range of stimuli. Recent work shows that we could have obtained a more nearly human fallibility from the model by employing a shorter time constant in the postdetection filter.

The fourth column is the percentage of agreement between the model and the average of the three observers. Note that the model's prediction of the subjects' responses

TABLE II.—*Comparison of the Electrical Model With Human Observers in a 2-Alternative, Forced-Choice Experiment*

E/N_o	$P(C)_m$	$P(C)_o$	$P(A)_{o,m}$	$P(A)_{o,o}$	$P(C)_{so}$	$P(A)_{m,so}$
12.8	98.8	92.6	92.4	88.9	99.4	99.1
7.6	91.3	80.4	80.8	76.6	94.1	94.4
5.7	87.3	76.2	77.6	75.2	91.1	93.2
4.8	85.4	72.8	73.6	70.3	87.4	89.8
4.0	79.3	67.2	71.8	70.1	78.7	86.4
3.1	77.0	65.9	72.0	66.5	78.6	88.3
2.5	71.3	60.1	66.1	60.7	68.9	81.7
0.0	50.0	50.0	65.0	63.0	50.0	81.9

is better than their percent correct. That is to say, the model predicts their response better than the presence of the signal does. When no signal occurs in either interval, the model predicts their responses 65 percent of the time.

The fifth column is the average percentage of agreement between one observer and the other two. When this column is compared with column 4, it is seen that the model predicts the responses of the human subjects better than they predict each other.

The sixth column is the percentage of correct responses made by the subjects to the stimuli on which all three agreed, whether right or wrong. The column shows that the percentage correct, for the stimuli on which the three observers agree, is considerably higher than their percentage correct for all of the stimuli. This is, of course, to be expected—the multiple observer is better than the single observer. The last column shows the agreement of the model with the subjects on those stimuli where the subjects all agree. Again, the percentages are higher. Even when no signal is present, the model agrees with the three subjects on more than 80 percent of the trials.

We may conclude that the model is apparently responding to the aspect of the stimulus most important in human signal detection, a considerably smoothed representation of the stimulus envelope.

A Mathematical Model of Monaural Detection

A brilliant paper by McGill in 1967 has shown that the results of an early experiment by Marill in 1964 can be explained in terms of an energy-detector model. Marill had employed an envelope detector in his derivations and had arrived at a formula for predicting the percentage of correct responses in a two-alternative, forced-choice experiment. McGill arrives at the same formula by way of an energy detector. He assumes that a narrow band of noise, or noise plus signal, is gated for a time T and the resulting voltage squared and then integrated. The integrator is discharged between observations. From the statistics of this device he derives Marill's equation.

McGill then goes on to show that the bandwidth assumptions made by Marill in fitting his theoretical function to human observers are inappropriate, and that a better adjustment can be made by assuming a different number of degrees of freedom in the probability functions. He shows that the Rayleigh-Rice statistics employed by Marill can be replaced, and more generality achieved, by employing the noncentral χ^2 distribution.

The electrical model we have been discussing can not only vote in a 2AFC experiment but, by recording samples of its output, can generate the distribution functions of its underlying statistics. If we sample the noise distributions measured at the output of the postdetection filter, we obtain a probability density function that resembles, but differs from, the Rayleigh distribution. It is less skewed, but still has considerable skewness. It does not resemble any χ^2 distribution. The resemblance to the Rayleigh

distribution suggests that the appropriate function would be a Rayleigh-like distribution with more degrees of freedom, and this proves to be a special case of the χ density function. Figure 12 shows a χ distribution with 14 degrees of freedom. The points represent 10 000 samples of the output of the postdetection filter.
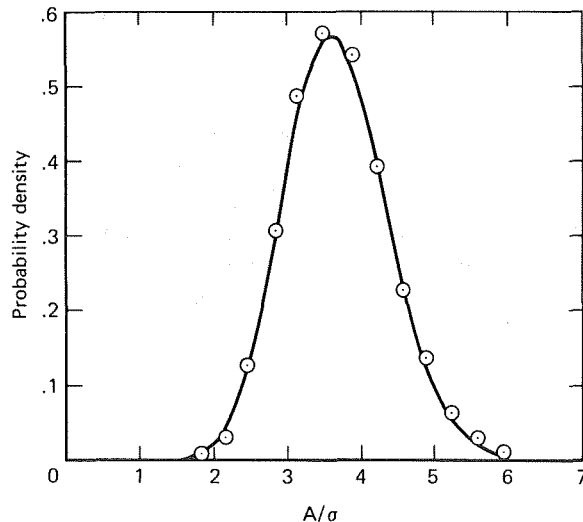
FIGURE 12.—χ distribution, $\nu = 7$ and data from electrical model.

Because the χ distribution fits the data for noise alone, the next question is whether the noncentral χ distribution, with the same number of degrees of freedom, will fit the data for noise plus signal. Figure 13 shows the resulting "psychometric" function. The abscissa is the signal-to-noise ratio and the ordinate is the difference of means divided by the standard deviation of the difference. The fit appears to justify the assumption about the appropriateness of the distribution functions.

Noncentral χ Distribution and Psychometric Data

Figure 14 shows the same χ distribution and another with 10 degrees of freedom along with data for Marill's two subjects. It will be seen that one of the subjects fits the curve for $\nu = 7$ (14 degrees of freedom) very well. The other subject apparently requires fewer degrees of freedom and even
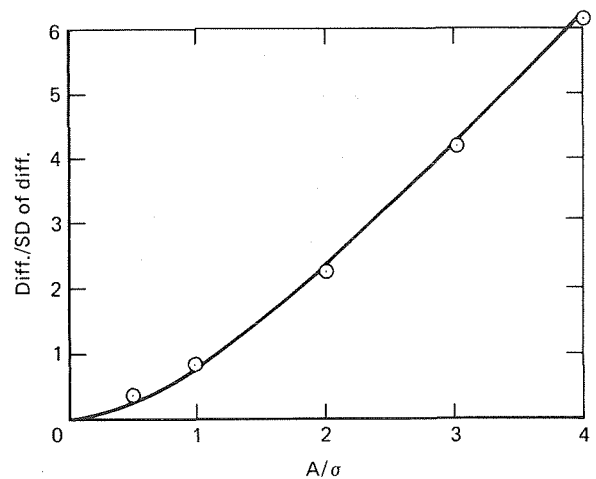
FIGURE 13.—Noncentral χ distribution and data from electrical model.
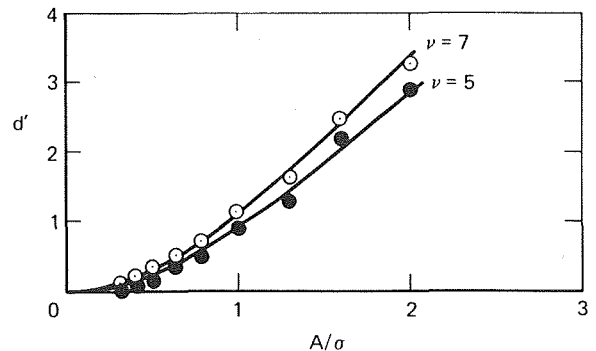
FIGURE 14.—Noncentral χ distribution and psychometric data.

then yields a rather ragged fit. Apparently, the parameters chosen for the electrical model (50-Hz bandwidth and a time constant for the postdetection filter of 50 msec) correspond reasonably well with the parameters employed by the first subject. The data for the second subject require the assumption that he employs either a wider filter (Marill's conclusion) or that his integration time is shorter. At the present state of our knowledge of individual differences, it is not possible to decide which (or both). The raggedness of the second subject's fit also suggests that nonstimulus factors are influencing his behavior, attention lapses, indecision about which button to press, and so forth.

The rather surprising agreement between the data for the model and for one of Marill's subjects suggests that this subject, like the model, is governed in his responses almost wholly by the statistics of the stimulus. The parameters chosen for the distribution employed are well within the range of values estimated for detection experiments—a "critical" bandwidth of 50 Hz and a time constant of 50 msec. The latter is the figure recently reported by Zwislocki for the auditory system.

Comments[1]

THEODORE G. BIRDSALL

University of Michigan

THEODORE G. BIRDSALL: I would like to talk about the problem of research and real solutions to real problems. I think I know what research is and I think I know what real problems are, because I have had a lot of exposure to them. But I have seen very few real solutions, even though research is supposed to give them to you.

There have been a lot of nice quotations concerning this business of research and real solutions. One that keeps coming back to my mind is from John Pierce, a research director at Bell Laboratories: "For every good research idea that comes out of basic research, it takes 100 good engineers to make it useful." That seems about the right proportion to me. Even when most people in the research atmosphere say they are finished with a problem, the solution is a very long way from doing anybody any good. We need to think about these 100 engineers. I do not think they are always in the right places, especially in connection with university-type research. They are nonexistent in a lot of places. In fact, I think we are missing what should be a whole profession.

The goal of this profession would be to connect some of the research ideas and the real problems and try to get some real solutions out of them. They should not be doing the research but they should be aware of its present state. When we train Ph.D.'s now, they imitate people who are doing research or they do research themselves in a field that has not been touched. This trains them to do new research. However, I am not looking for people who will do more research. I am looking for people who can understand it, and understand very thoroughly all the things that it is not. They must be able to do this with a great many pieces of research and then look at the real problems with all their nasty particulars and try to get the two of them together.

There is also product-oriented research. Most of the companies that I am familiar with care only about the product that is going to come out of it and how the product will affect sales and profit.

I am not going to break research down into basic research and applied research. That is a very dangerous cut to make. I am talking about both. But there is certainly what we call mission-oriented research. As a Navy contractor, I am reminded repeatedly that my mission is to make that fleet stay above the water, or below the water, whichever it is supposed to do. Whatever I may want to do in signal detectability, the main object is the particular mission.

We also have educational institutions that are interested in knowledge-oriented research. Again, I want to emphasize that I am not distinguishing between basic research and applied research even in this knowledge-oriented research. Often we go after certain kinds of knowledge because we know that it is necessary if we are to accom-

---

[1] Mr. Birdsall commented on and led a general discussion of problems associated with applying the results of scientific research to the real world.

plish some goal. But first we have to get the knowledge. The same kinds of people man all three categories of research. We do not have people in between who are selecting various items of knowledge research to put into actual use. Of course, there is always the pressure to just "get the job done," although the pressure is not so much felt in the universities. But this applications work needs a lot of attention and I think the prob-

lem is more than just a lack of communication or information. If the problem were only a lack of communication, it might be handled by something like publications. But I think that a lot of hard work is needed if we are going to continue converting partial results of research—and almost all the results are partial results—into some sort of practical engineering theory and then into some sort of engineering practice.

DISCUSSION

LLOYD A. JEFFRESS: We know a couple of things as a result of knowledge-oriented research that I have tried to tie up to mission-oriented research.

One is this 10- to 15-dB improvement in the detection you get with two ears over one. The other is the remarkable ability of the ear to detect differences of direction. Every attempt to realize 15 dB in practice has just not produced anything. I keep wondering about localization. Maybe this has some practical applications. A pilot might be able to use his ears to determine bearings instead of the eyes when his eyes are busy looking at instruments. It seems to me to have considerable promise, but it would take quite a bit of engineering to realize that.

BIRDSALL: That is part of the point I am trying to make. It does take considerable engineering. It is not trivial work, and it is very hard work. With the people I have been associated with in this type of research the typical problem goes like this. They have a piece of gear that was designed 8 years ago and it does not work today. It will be with them for the next 5 years. It must work tomorrow, and they want answers now. They do not have time to look at some abstract idea that you have which looks very promising. It is not something that they can assign to a man and have him think about for a while. It is something that might require the work of a team for a year, and then they might come to the conclusion that there is no conceivable use for this kind of an idea at this time. We do not seem to put our money on anything until we have a large task problem that must be solved. There are occasional cases where somebody is sufficiently personable or arrogant to do everything himself. He gets out of the research climate and becomes a manufacturer. This is sort of an anarchistic way of getting things done, although some people have claimed that the best way to handle a good idea is to take it all the way from the research through the development: through the manufacturing, through the prototype trials, and all the way to the final field test trial. That is the only way to get the job done. It seems like a waste to have a research man do all that, because he is probably particularly good at only one of those tasks.

JOHN A. SWETS: How could you change the payoff structure in order to turn out people who would spend their lives being acquainted with research and with actual problems and converting one into the other? I agree that we are missing that profession of people. Is there any chance of bringing about that profession, and how would you bring it about?

BIRDSALL: I think the biggest motivator for it would be to have it recognized as a profession in the sense that there are jobs, and people doing this kind of thing. At Michigan we have a so-called professional degree in engineering. It is not a very honorable degree and does not confer the title "doctor." You have to study just as hard and do everything about the same, except you do not have to pass both languages. Also, the thesis does not have to be an original piece of research. Usually the student takes a doctoral thesis on something he has been working with and turns it into a practical system. It is a very dishonorable degree. Very few people go for it, because it does not have the word "Ph.D." tacked on it and the persons capable of finishing this degree are capable of getting a Ph.D. The payoff is not the same.

I think that what is most needed is a profession, a recognized body of people doing such work. Not doing research but getting the results of research into use. It is done in the large companies. Certainly Bell Labs puts pressure on their research people to first save the company a million dollars, then they can go off and play with what they want. But first they have to do something useful and profitable. So the people coming in from the bottom get into this work force. Their objective is not to make a name for themselves, but rather to do something that really helps the company. That is the name of the game and it works quite well. The company takes their research results and starts getting them into the practical system. For example, someone doing research in pattern recognition designs a little writing board that a very sloppy telephone dialist can use to write the number across and which then dials the telephone. These things get used.

This is something that I feel there is a great lack

of in smaller groups and in Government contracting. The contractor usually has the burden of making the research useful. But this is not the kind of thing you can do by direction; it takes a lot of effort.

JOSEPH MARKOWITZ: You said it might be a waste of a person's time to take an idea and follow it through all the way to product development, perhaps to sales, because some people are better suited to some of these occupations than to others. And yet you treat your two degrees as if the qualifications for one were identical with the qualifications for the other. Is there not a way to emphasize the difference?

Maybe you should make both programs much more difficult, so as to take advantage of the inherent specialties that you propose are really in people.

BIRDSALL: I think I was mixing two kinds of points there.

MARKOWITZ: Alternatively, is that a way to get people into this field you want?

BIRDSALL: The majority of people who are in research are not always the best ones to put into research applications. There are occasional ones who can do both. But the people who are going for engineering Ph.D.'s are pretty much oriented toward research. Why? That is what our business is. We do a lot of research. You have got to have a thesis to graduate. Besides, I do not think you really want people who are very good at research to be doing research applications. They may not be very practical. Whatever makes a person imaginative and causes way-out ideas to come into his head may be the same thing that prevents him from thinking logically and in cold, hard terms.

SWETS: But universities hire those kinds of people for their faculties. It is a little hard to understand how those kinds of people are going to turn out practical people. If you are looking for a man who will go around finding out where research is and applying it, the chances are you will not find him in a graduate program at a university. And the chances are that he should not be there. He ought to be where he can get some decent practical training, rather than the kind that the university will give him.

BIRDSALL: The Ph.D.'s are pretty much aimed toward research, and that is not a prerequisite for practical work.

MARKOWITZ: Would you propose some kind of an apprenticeship?

BIRDSALL: Primarily, I propose a job. I do not think we lack for people who are capable of doing this kind of work. My own experience, which is very limited, is with the Navy laboratories. We are trying to make university research more useful. What problems are Navy laboratories saddled with? They either have prototype equipment that is going into the fleet as soon as possible or they have equipment that does not work.

STEVEN E. BELSLEY: I want to know why the Navy thinks it has to turn to the universities for this kind of a service. It seems to me that this is just what a contractor is set up to do. The product people are supposed to produce equipment to perform certain missions, and they draw upon a broad spectrum of resources to do this, including the products of university research. If you can divide research into your three categories and assume they never overlap. . . .

BIRDSALL: I am not worried about their overlapping. I am worried about the large empty set in the middle. That is what I am trying to fill. For example, one of these research outfits predicts a 15-dB improvement in certain circumstances. What does that have to do with practical detection situations where it might be useful? This man over here has a piece of gear that must work every place in the world under all sorts of different conditions. He does not have a year to determine whether it is worth the extra weight and cost to use it. He does not know if there will be the appropriate conditions under which he can get the 15-dB improvement. He looks at it and he says, "Forget it."

BELSLEY: It seems to me that you are talking about mission-oriented research, which is the proper subject of a Government laboratory operated for that purpose. The product-oriented research is the proper subject of a manufacturer. Knowledge-oriented research is the proper subject for the university.

BIRDSALL: And Government laboratories also.

BELSLEY: They participate in knowledge-oriented research to some degree, but that is not their prime object in life. Their prime object should be mission-oriented research. Between 1950 and 1955 there was a flow of prominent people from the Government laboratories into industry to do product-oriented research. In order to fill the gap, Government laboratories have turned to the universities.

BIRDSALL: If you look through the ideal industrial structure, it goes all the way from research through manufacturing and sales, and into profit. There is a nice orderly progression to workable ideas. What I am trying to say is that there is something missing. There is a gap. It is a long way from research to advanced development. For advanced development you almost have to have a block diagram and the circuitry you are going to build.

BELSLEY: That is the way the Department of Defense runs its railroad. That does not necessarily mean it is the right way to do it.

BIRDSALL: That is the way it is being done.

BELSLEY: Not always, but there are other ways of doing business. Advanced development is so much dependent on the goal that has been set by the customer. It is heavily structured within that framework. You have to consider that there are certain things structured within the desires of the customer. You get an idea that you think could be applied to one of his burning problems and then you start to develop along that line. Then you bring the idea up to a given state so the customer can make a decision as to

whether he wants to proceed. When you are trying to sell the idea that you should do research and you are interested, from the research standpoint, in how somebody could use this research, you should structure research with the customer in mind. If you are going to fiddle for fiddling's sake, you belong in education.

BIRDSALL: I agree with you, but I do not think most of the people in research—in advanced development—are the ones best suited to help make the tie between research and applications. Often the people in research make up little stories and convince themselves that they are doing something useful. They do not have much contact with the real world. We need persons who are not spending all their time doing research and who are not spending all their time putting together hardware, but who are trying to get better cooperation between research and advanced development.

MARKOWITZ: Is there really much intrinsic reward in the task?

BIRDSALL: It is fantastic. I would gladly join them.

MARKOWITZ: Then why are there no people?

DOUWE B. YNTEMA: Let me answer that. I think you are missing the crucial point. It is true that you have to find a group of people who have or will be given the right motivation and the right training. I suspect that you are also quite right that the system is not developing people who will be satisfied with this type of job. This is a problem, but I think the big problem is organizational. You said earlier that as many as 100 such people might be needed for one research idea. This work has to be done by large organizations, like Bell Labs. These organizations are large enough to have the depth and resources that enable them to build a team of 10 or 20 men. When their job is done 5 years later, the men become members of other teams.

BIRDSALL: Most of the people I know in advanced development are given a mission for a specific piece of gear. They just do not look at anything that is not obviously related to that piece of gear. Somebody has to tie together several pieces of research and make them into a working system that these people can understand and almost prejudge for its potential usefulness.

BELSLEY: Is that not an area where the systems research analyst fits neatly? He considers a broad spectrum of possible applications in the broad spectrum of ways that information can be fed into the system. He then tries to put it together into a conceptual pattern that will meet these goals—all without having to build any equipment. These are what are called feasibility studies.

BIRDSALL: You must know some good systems people.

MARKOWITZ: It seems to me the situation is sort of hopeless. We do not have enough people because,

in the long run, every research idea has got to take up a large number of people to keep track of it as well as the mission-oriented problems.

BIRDSALL: How long does it take to ratify a research idea? Four, five, six years? But somebody can read 18 months of research in 18 minutes.

MARKOWITZ: Are you now saying that it does not take as many as 100 applications people per researcher?

BIRDSALL: That is right. You need a large number of them compared with the number of researchers, but, for mission-oriented research, the ratio may only be 8 to 1.

BELSLEY: I think what you are calling mission-oriented research is what Bell Labs calls systems research.

BIRDSALL: A lot of the young people doing so-called research at Bell are in very directed research. Their objective is to look throughout the whole telephone system and see where certain ideas can be put to use.

BELSLEY: They are not doing research. They are seeing where they can apply what they have learned. Once they have got an application, then Bell says, "You turned out to be a good applications man, now do some research."

BIRDSALL: One out of eight passes from systems research to basic resarch.

SWETS: So we can identify 100 people at Bell who are doing applications research, and 50 people at Lincoln, or maybe 500. That is not very many. There ought to be other groups besides Lincoln and Bell. There are a few systems analysts in companies around, but I do not see them doing very much of what we are talking about.

I am also worried about another thing: I do not think that the universities are going to build these people with strange degrees. The question comes to mind: How might an agency with a mission build some people to fill the need?

JEFFRESS: At the present, if you work on a military problem, you do so at the cost of your own career, so to speak. But a few years ago, the Defense Department was offering overtime pay to faculty members who would devote some of their time to these problems. It seems to me that this encouraged a lot of intelligent people in the academic field to work on such problems.

SWETS: To work on military problems of interest to them. But they were not really looking at what research was coming out and what problems existed?

JEFFRESS: No; that is true. But it was a closer bridge than we have got now.

YNTEMA: What do you think of the institutional pattern the medical people have set up? Would you think that their ratio is more nearly correct and that their institutional structure is more nearly correct?

BELSLEY: Recently, the medical profession has

been criticized for spending too much money and time gathering information, and not enough applying this information to medical problems per se. The main complaint is that for 10 years we have been dumping a billion dollars a year into the medical research field. How is this being used to prolong life, to cut down strokes, to cut down heart attacks, and so forth? As a result of these complaints, the medical profession is gradually orienting its research organization toward thinking in terms of applying all this information. They are not saying that they intend to cut off medical research per se, but they have set up a new institute that is essentially mission-oriented.

R. MARK PATTON: The profession Birdsall is suggesting seems to me not so different from the ideal concept of human-factors engineers within an industry. I emphasize "ideal," because rarely do such people have sufficient impact on the system development process. I spent several of the most frustrating years of my life as a human-factors engineer, and finally came to the conclusion that the job was all hopeless. You are always operating in the context of a very complex tradeoff situation and the human-factors requirements often have to take second place to weight, volume, power, and cost. The earlier phases of system development, where future systems are being outlined are just as bad. The tradeoffs are still there, and deadlines are such that you usually have a maximum of about 3 days to respond to the flap of the moment. I feel that human-factors engineering ought to work somehow—I have no idea how—to bring the best available research knowledge to bear in some very realistic way on the development of a system, but there are many obstacles.

BIRDSALL: That is why I say there has to be a profession in which persons have a chance to sort out the research that should go with some development before the flap comes along, because when that flap comes, you only have about 3 days.

EARL A. ALLUISI: There is an easy way to get such a profession in our capitalistic society. Create the positions and attach salaries to them. Pretty soon you will have people going into that line of work.

PATTON: There is also the question: Is the academic Ph.D. really the proper training for an industrial human-factors person? If you look at the roster of the Human Factors Society, a typical training is either a master's or doctorate in psychology. Sometime ago in *The Human Factors Society Bulletin*,

there was a list of the places that even pretend to offer a specialized training in human factors. The list was only one page long. I think maybe that is part of the problem.

ALLUISI: I have argued for 15 years that to put us at the board doing human-factors engineering is a misuse of Ph.D. training in psychology. That is engineering; it is not what a Ph.D. in psychology is aimed to train a person best to do. There should be a specialty in the engineering school where psychology is taught and its application is taught—just as in mechanical or civil engineering. I can name at least a half a dozen engineering schools that are now beginning human-factors-engineering programs as a basic engineering curriculum. I suspect this will develop further, and I believe that in a few more years we will have more people doing applications, provided the market remains one that encourages them to get into it. We train Ph.D.'s to be researchers, and we try to make them knowledge oriented. Knowledge orientation may be mission-knowledge orientation or even product-knowledge orientation, but a Ph.D.'s orientation is going to be slightly different from what you want in applications. It is going to be just as different as that expected between a physicist and an engineer. You cannot easily get a physicist to do engineering.

MARKOWITZ: Assuming that we can create the position and attach sufficient monetary gains to it to attract the right people, it seems to me there is still a problem. I can tell these persons where to look to find out about basic research. I tell them what journals to look at and teach them how to read journal articles. I can identify good people in the field to help them sift through the material. I can get them fairly well acquainted with current research effort. I would like to know how to acquaint them with mission-oriented problems.

JOHN W. SENDERS: Was there not once a Government publication of things that the Government would like to see invented?

BIRDSALL: I do not think there is as much problem getting people acquainted with mission orientation. That is where the pressure is. The pressure is on the people who have the problems. I personally feel you would get ten times the usefulness out of research if you take half the people in research and put them together with some other good people who are mission oriented. I favor making the research useful instead of just making research good.

Application of Decision Theory to Manual Control

JEROME I. ELKIND

Bolt, Beranek & Newman, Inc.

In the last couple of days, we seem to have progressed from some very specific problems to some very general ones. I want to return to some specific simple problems and to some of the issues that Rathert and Yntema raised yesterday. I am still dissatisfied with the taxonomies that have been presented, largely for the reasons that Birdsall just mentioned: from the theoretical point of view, the various kinds of decisions that we have been talking about look the same. Thus, are they really different, and if so, in what respects?

To illustrate this similarity of decisions, I will spend a few minutes talking about what Rathert called reflex kinds of decisions or what Yntema spoke of as decisions that are made by satellite computers. Let us consider some very simple manual-control problems in which a decision theory framework is appropriate. These are control problems in which the reflex kinds of activities are dominant, and in which the basic behavior that we are interested in looks as though it can be modeled nicely by the kinds of models that are used for some of these more complex decisions. First I will describe the kinds of control situations that we have investigated, then I will describe briefly an experiment and, without going into too much detail, will discuss some of the results. Finally, I want to say a few words about the direction of current work in manual control. This work is based heavily upon optimal-control ideas, which I think tie in nicely with the kind of decision theory that we

have been discussing today and yesterday.

In the experimental situation with which we have worked, there was a random gaussian type of input signal with low bandwidth. This input is compared with the response of the system and an error signal is derived. The human operator looks at the error and makes an appropriate control movement that is fed to a dynamical system. The system, in turn, responds to this input. This is the classical kind of feedback-control system. Our experimental situation differed from that used in most previous work in that the dynamics of the system were time varying. The human operator had to maintain control of the system in spite of the variations in dynamics. An extreme example of this kind of control situation would be the Eastern Air Lines plane that, when flying from Boston to New York recently, encountered another aircraft, lost much of a wing, but landed satisfactorily.

In our experiments, at some arbitrary, random time the vehicle's dynamics suddenly changed. When a change in dynamics (which we call a transition) occurs, the subject must first detect the fact that the dynamics have changed, he must identify the new dynamics, and then he must adopt a new control strategy appropriate for the new dynamics. This whole decision process must be accomplished in a very short time to prevent the system from reaching a nonrecoverable state.

We had four possible sets of dynamics that could be operating at any one time. The one that the subject was tracking was called

$C_0(s)$. Maybe once every 15 seconds during the course of a 4- or 5-minute tracking run these dynamics could change. Any given dynamics could change to any one of two others. We gave the subject a button that he was to release when he had detected a change. This enabled us to distinguish the subject's detection of the change in vehicle dynamics from the subject's identification of the new dynamics and modification of his control status.

Figure 1 shows the time history of a typical transition. The input-forcing function and the system response are superimposed on each other. The control movements that the subject made and the error signal are shown separately.
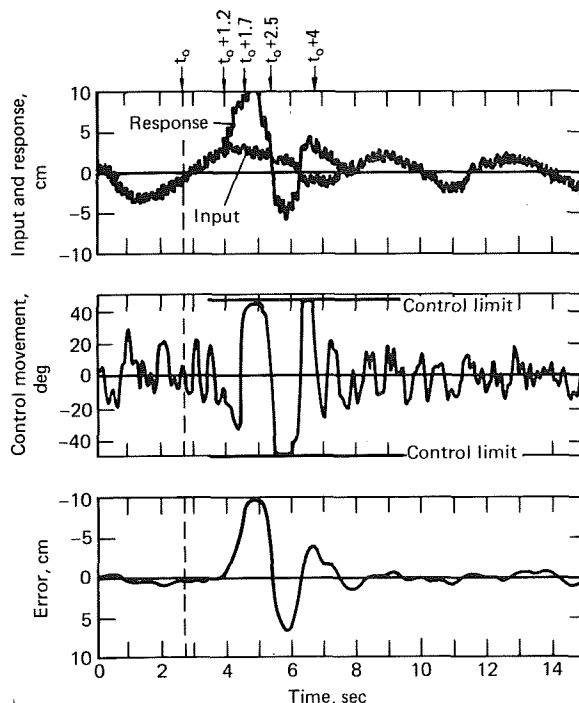

FIGURE 1.—The time history of a typical transition. Input-forcing functions and system response are superimposed.

The actual change in dynamics occurred at the point marked t_0. For this run, the dynamics were K/s^2. That is, the system behaved like a pure inertial system and the subject controlled the acceleration applied to the system. The transition in this run was one in which the polarity of the dynamics changed and the gain of the dynamics increased.[1] Figure 1 shows that the subject's behavior remained relatively unchanged for about 1 second after the transition, in spite of the fact that the system was highly divergent. During this time, the error remained small. About 1 second after the transition, the error suddenly started to increase very rapidly. The subject apparently detected a transition; shortly after this he made some fairly violent movements to try to cancel the rapid increase of the error. During this period, he obviously modified his characteristics and must have completed an identification of the new dynamics.

About 4 seconds after the transition, the subject managed to get the system more or less under control and the error settled down. The subject's movements were more or less of the same character as they were before the transition. He had canceled out the initial transient, but it took a little while before his movements settled down to an appropriate amplitude. Notice that the posttransition movements tend to be smaller than the pretransition; an effect that is accounted for by the increase in the gain of the dynamics.

You can play a lot with these kinds of records. We have studied the control characteristics as well as the things that are pertinent to the decision process itself. But today I want to talk about the decision parts of the process, mainly those that we call detection and identification: detection of the transition and identification of the new dynamics.

Once a subject detects a transition, he is able to make very rapid, gross changes in his control strategy. This is illustrated in figure 2, which shows polarity reversal of transition with K/s (velocity) dynamics. The gain of the human operator's describing

[1] By polarity change we mean that the movements that the subject was making become inappropriate in the sense that they were in the wrong direction, resulting in an unstable system. By gain increase, we mean that the system amplifies the subject's control movements more than before. This can also lead to instabilities.
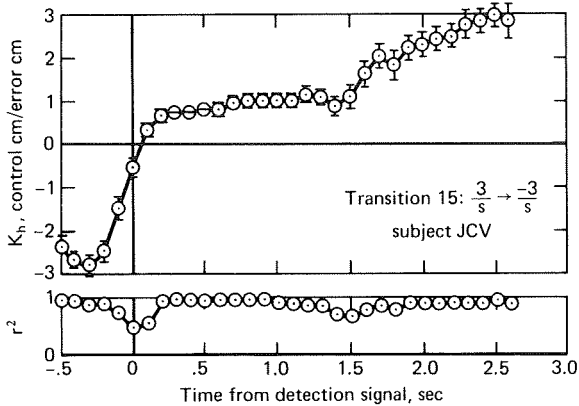
FIGURE 2.—The polarity reversal of transition (time from detection signal versus K_h).

function is plotted in figure 2 as a function of time from the release of the detection button. This kind of transition usually occurred about 0.4 second prior to the release of the detection button. In about 0.4 second, the subject has changed the polarity of his gain. However, his movements are a little low in amplitude as he cancels out the transient. After a couple of seconds, he starts to increase his gain gradually until it finally has approximately the same value as it was before the transition, except polarity is reversed.

We observed that the whole adaptive process happens very much at the subconscious level after subjects gather experience with the experimental situation. If they did not have the detection button to release, subjects would often not be aware that a transition had occurred and that they had changed their control strategy. They would just go ahead and do it. So here is a case where, in Yntema's terms, the whole pattern recognition process and the selection of a new computer program is at the unconscious level.

STEVEN E. BELSLEY: What sort of a device is the subject operating?

ELKIND: For the experiments that I am discussing, there were two kinds of dynamics: K/s, pure velocity control, or K/s^2, pure inertial control. Physically, the subject is sitting in a room with a scope and a joystick in front of him.

BELSLEY: Is it an overdamped system?

ELKIND: It is just pure integrator. We have done some work with overdamped and with unstable systems of various kinds, but none of it was done systematically enough to present.

I think this is an essential point to be made: Most of us have been talking about models that we think have the right structure for representation of human behavior in complex tasks. In most of the cases that have been reported, these models have been applied to very simple situations only. I think the importance of these experiments lies not so much in the results as in the structure of the models.

Now consider the problem of detection. For this kind of problem, detection is expressed very simply. You might argue that the breakdown into detection and identification, which are two separate problems, is somewhat artificial. But we do have the detection button, which presumably indicates detection, so it is worthwhile looking at. There are two probability measures that the subject is going to be concerned with: one is the probability, based upon his observations of the error signal and his knowledge of the stick movements he is making, that the initial dynamics are C_0; the other is the probability that they are not. We postulate that the subject makes estimates of these probabilities. Whether or not the subject wants to push the detection button depends, of course, on such things as his utility functions, and so forth. The tracking problem is very much one of sequential decision-making. Things are going on all the time. When the polarity of the dynamics change, not only does the subject have the opportunity of making repeated observations, but he can tell from the signals if he is doing something wrong. The result is a complex interaction between what the subject is doing and what he is seeing.

Rather than try to work with continuous signals, we have quantized them and assumed that the subject makes observations, say, every 0.2 second, which corresponds, roughly, to the time it takes for him to make a movement. Under these assump-

tions, the subject is viewed as operating on sample values of the signal, which allows us to know, more or less, what data he is observing.

There are two components to the data: the error and the control movement. The subject takes samples of the error signal and also the control movement, which are the only two kinds of information available. In a real airplane where there are motion cues and other information, the system response would be much more complex. Note that the error is somewhat complex, because in general the subject must worry about the error magnitude, the first derivative, and so forth. Thus, the error state is a multidimensional quantity. Actually, we need to be concerned with only those properties of the error that the subject can perceive. In these simple experiments, we were able to simplify things considerably on the basis of some earlier work that we had done. Most of the detection and identification of behavior can be well predicted from the change in error rate from one sample to another.

Our model looks only at the subject's change-in-error rate from sample to sample and the control movements that he makes during a sample. There are some control-theoretic reasons for concentrating on error rate, but I will not go into them at this point. Suffice it to say that the model seems to work fairly well, and there are reasons why it should.

The next step is clear. We want to use Bayes' rule to devise the posterior probability that the dynamics are still C_0 after the subject has made each observation. We express posterior probability as

$$P(C_0|\Delta e, c; n) = \frac{P(\Delta e|C_0, c; n) \ P(C_0; n)}{P(\Delta e|c; n)}$$

where Δe is the change in error rate and c is the control movement during the nth interval. Keep in mind that we have the sequential situation here. The prior probability at the beginning or the nth control interval, $P(C_0; n)$, is derived from the posterior probability at the end of the previous interval.

One can derive a similar expression for the posterior probability where the dynamics are not C_0. However, because this is a sequential problem, most of the behavior of the model is going to be determined by the conditional probabilities $P(\Delta e|C_0, c; n)$ and not by the initial assumptions made about the prior probability of C_0.

We are actually concerned here with subjective probabilities, with the subject's perception of the conditional distribution that a particular error Δe would be observed if the dynamics were C_0 and if he made the control movements that he did. There are several sources of error that affect these probabilities. First, the subject does not know the input signal. During the control interval, the input signal could have changed its rate appreciably and contributed to the variance of the subject's estimate. Second, the subject does not always make the control movements that he thought he made. Third, the subject does not always see the real error. So we have some distribution of the Δe.

We make the following assumption, which I will justify briefly. All input signals to the system are gaussian. We also assume that the subject makes an unbiased estimate of the mean of that distribution. That is to say, the mean of the conditional distribution is exactly what the change in error rate ought to be, if, indeed, the subject made the control movement c, and if there were no input disturbance. Then we would go on to say that the variance of this distribution will have two components: one that is the variance of the input signal, the other that is a proportional to the mean Δe. If the subject makes a larger movement, there will be a larger change in error rate. We would expect that the subject's estimate of change would show larger variance. This is a case of a constant proportional error.

We have ways of estimating, or at least placing reasonable bounds on, the various components of the variance. We simulated this model on the computer and exercised it with the actual data used in the experiment. I would now like to compare the experimental results with the model results.

In figure 3 is plotted detection signal times (the time after the transition at which the subject released the button) against the detection time predicted from the model. Data are shown for three of the six transitions we worked with. The points are scattered, as would be expected, but they fall along the line of unity slope, intersecting the origin at 0.4 second. The intercept, 0.4 second, is the effective reaction time to this kind of transition in this particular kind of experiment.

FIGURE 4.—Posterior probabilities computed from the model as a function of time from the transition.

FIGURE 3.—Detection signal times versus the predicted times.

The next problem is one of identification, in which the subject is concerned with more than just whether the dynamics are C_0 or not. He must determine which of the three possible dynamics are in effect: C_0, C_1, or C_2. If one is willing to assume equal values for all outcomes, then the subject should decide in favor of the dynamics whose posterior probability is largest.

Figure 4 shows how the posterior probabilities computed from the model change as a function of time from the transition. The initial dynamics in this case is C_0 and C_1 is

the correct choice that the subject should have made. The dynamics actually change from C_0 to C_1; the alternative dynamics are C_2. The transition occurred at the point marked t_o. We find that the probability estimate of C_0 drops, starting at the transition time, and gets below about 0.5 at about 0.4 second. The probability that the dynamics are C_1 rises above 0.5 at about the same time. In this case, all of the probability estimates of C_2 remain small. According to our detection model, the subject should have released the detection button at about 0.4 second after the probability of C_0 became less than 0.5.

We will now consider the identification performance of the model and the subjects. In figure 5, I have plotted results from four transitions: (a) gain increase, (b) gain decrease, (c) polarity reversal, and (d) another gain decrease. The probability of the correct dynamics, $P(C_1)$, are plotted against the probability of the incorrect dynamics, $P(C_2)$. The probability estimates are those generated for the first time interval that the model said that identification should have occurred. The black points indicate that the subject correctly identified the transition, and the open points indicate that the subject incorrectly identified the transition. If the points lie above the diagonal line, $P(C_1)$ is
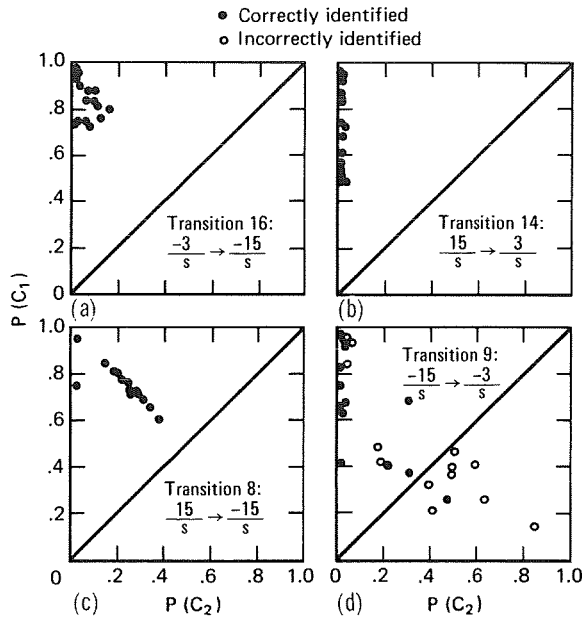
FIGURE 5.—Results from four transitions: (*a*) gain increase; (*b*) gain decrease; (*c*) polarity reversal; (*d*) gain decrease. Probability of the correct dynamics $P(C_1)$ are plotted against the probability of the incorrect dynamics $P(C_2)$.

greater than $P(C_2)$ and the model would have correctly identified the transition.

Let us consider the transition (fig. 5(*d*)) in which many of the points lie below the diagonal line. Most of these points are open, meaning that both the model and the subject would have incorrectly identified the dynamics. In this transition, there was a gain decrease that the subjects mistakenly took to be a polarity reversal. This kind of error is not uncommon and is easily explained. When a gain decrease occurs, the control becomes more sluggish than it was before, so that the same control movement has much less effect on the output. If the input signal is moving fairly rapidly at the time of transition, the error will increase with every movement that the subject makes. This behavior resembles that obtained with a polarity reversal, where every movement by the subject typically results in a larger error. The errors shown in figure 5(*d*) are, therefore, a result of the coincidence of a rapidly changing input signal and a gain decrease transition.

BELSLEY: The subject would not necessarily make this error if he were tracking on zero error.

ELKIND: He was tracking with the error near zero. This was not a terribly hard tracking task. If the system were essentially quiescent, so that the subject was not required to make any movement, he would not get much information upon which to make a decision. Presumably, he would stick to the initial dynamics. Only when the subject really starts making movements can he get information from the system.

About 13 of the transitions were incorrectly identified by the subject, and 8 of those were also incorrectly identified by the model. Actually one of the transitions that the model identified incorrectly, the subject identified correctly. But we nevertheless get surprisingly good matches between model and subject behavior in terms of the number of incorrect identifications.

We checked the same model with the same parameters on another set of dynamics, K/s^2. Figure 6 compares the predicted and observed detection times. We did not have a detection button in these experiments so we had to estimate when we thought detection
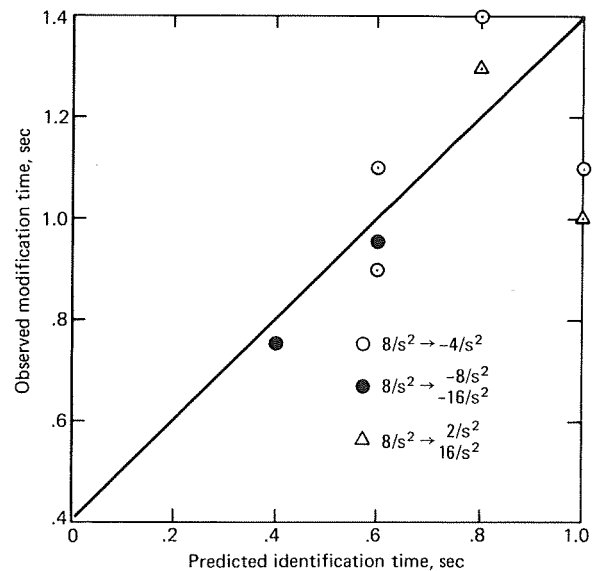

FIGURE 6.—Comparison of the predicted and observed detection times.

had occurred from looking at the subject's control behavior.

Table I illustrates the identification performance of the model with these dynamics and the posterior probability of each of the possible posttransition dynamics. Again, the same model seems to work pretty well. Occasionally, the dynamics having the highest probability would have a gain differing by a factor of 2 from the correct gain. For example, when the transition was $4/s^2$, the model might declare it to be $2/s^2$. This is not a serious error.

I would like to comment about conditional probabilities $P(\Delta\dot{e}|C_i, c; n)$ that play such a central role in the model. The interpretation of these probabilities is relevant to some of the questions that Rathert raised yesterday and perhaps also to some of the problems to which Belsley has been alluding. One interpretation of this conditional probability is that it indicates the extent to which the subjects, who in this case were well trained, had developed good internal models of the dynamics of the systems that they were controlling. The subject's state of training is embodied in this conditional probability and, more specifically, in the variance of the probability density. If the subjects are well trained, the variance should be small. Presumably, they can adapt very rapidly and will not make very many mistakes.

If, on the other hand, the model is not very well developed, and it has a large variance and maybe even the wrong mean, then the subjects will need more information to make the detection and identification decisions. They will take a longer time to get the additional data and they will be more likely to misidentify. So their state of training is embodied in this conditional-probability term.

Of course, the prior probabilities, or maybe the probability of a transition, come into the model in an important way. It ought to be related to the probability that a particular failure can or cannot occur. Perhaps one of the reasons that you see longer times for adaptation in actual flight situations than in laboratory situations is that failures do not happen very often. Thus the probability that a change will occur is low and it takes more data before the subject or the pilot decides that something has indeed happened.

In terms of data, none of these results solves practical problems. I would argue, however, that the model provides an appropriate structure for looking at these problems associated with system failures. It would be interesting to take Rathert's simulator, run some real dynamics on it, and try this approach in a fairly realistic problem. My guess is that it would yield reasonable results.

BELSLEY: Are you talking about a fixed- or moving-based simulator? You modify the problem by the introduction of motion that you have not taken into account.

ELKIND: I have not taken motion into account because I do not have it. But, in terms of the structure of the model, motion comes in very nicely. Take some quantity X

TABLE I.—$P(C_i)K/s^2$ *Transitions for Identification* C_i

Transition		$C_i(s)$							
		$-16/s^2$	$-8/s^2$	$-4/s^2$	$-2/s^2$	$2/s^2$	$4/s^2$	$8/s^2$	$16/s^2$
$8/s^2$	$-4/s^2$	0.04	0.33	[0.35]	0.22	0.03	0.006	0.02	0
	$-4/s^2$	.002	.04	[.29]	.44	.10	.02	.09	.0001
	$-4/s^2$	.02	.19	[.31]	.25	.06	.02	.16	.0003
	$-4/s^2$	.03	.13	[.43]	.12	.06	.03	.003	.0009
$8/s^2$	$-8/s^2$	.65	[.20]	.06	.03	.006	.003	.05	0
$8/s^2$	$-16/s^2$	[.75]	.19	.04	.01	.002	.0005	.006	0
$8/s^2$	$16/s^2$	0	0	0	0	0	.0001	.36	[.64]
$8/s^2$	$2/s^2$	0	0	.02	.12	[.56]	.17	.13	0

that is multidimensional and represents the set of things that the subject can perceive. These may be motion cues as well as lots of other attributes of the system, and just the error. Whether an experimenter can deal with all those things is a real question. That is essentially the question that Edwards raised in connection with a real situation. The complexity becomes enormous and it is difficult to do good experiments.

YNTEMA: You have got a nice line of defense in that the variance increases if the subject is finding the situation too complex to perceive very exactly. You have a parameter already sitting there.

ELKIND: That is the parameter of the subject's response to complexity. The question is the experimenter's response to complexity, and that is another question. This is very much like most of the other decision processes we have been talking about.

Let me try to draw a couple of more parallels between the kinds of things that are going on in manual control and those in decision theory. There is one theoretical framework with which you can build a theory of complex manual control systems—optimal or modern control theory. It is a theory that is designed to handle complex control problems. If one is going to worry about complex situations, it seems to me that one wants to invoke this theory which was designed to handle such situations.

The way in which problems in optimal control theory get formulated is as follows: We have some vehicle dynamics with disturbances driving it, and the state of the vehicle is represented by x. We have some displays that present to the pilot some transformation of x, which we call y_d, a vector.

There is some perceptual process that the pilot employs to look at these displays and to perceive y_d. He derives from that perception some variables y, which is what he thinks the displays are telling him. From them he must try to reconstruct the real state of the system, because that is what he is really interested in if he wants to control it. What he must do is to derive some estimate of the true state of the system, which goes to

the controller and produces some control action on the pilot's part. The control action is then passed on to the vehicle. In most airplanes, the instruments are spread around the panel and there is a visual sampling process that results from the fact that the pilot generally moves his eyes around.

One way of representing this perceptual process is to say that the pilot derives the displayed state of the system, delayed by some amount, and corrupted by noise. The noise represents the error in the pilot's reading of the instruments. But the variance of this noise depends upon where the pilot is looking. If he is looking directly at a display, then the variance would be smaller than if he is viewing it peripherally. Thus, the pilot's estimates of the state of the system are a function of the sampling strategy by virtue of the hypothesis that the noise associated with instrument observation is a function of the sampling strategy.

One of the nice things about modern control theory is that it explicitly states what the pilot is trying to achieve by the control process. To solve for the controller and estimator characteristics, we start by writing down a performance functional. The performance functional can have a variety of forms, but one reasonable form is a weighted sum of the mean-squared error, and the mean-squared control movements. The cost of sampling of moving the eyes can also be introduced into the cost functional. The computational problems involved in solving the optimization problems are not simple. However, the problem can often be segmented so that the controller and estimator characteristics can be found separately. If the cost functional is quadratic, the estimator will have the form of a Kalman filter. An intrinsic part of such an estimator is a model of the system being controlled. This ties in with the decision-theory model in which the state of the subject's training was represented by a model of the system being controlled.

Finally, one other small point. We have done simple preliminary experiments to look at the subjective cost functional issue. Al-

though a subject may be told to minimize some cost functional, he may, in fact, have some other (that is, a subjective) cost functional in mind. We have done some experiments with single-axis compensatory control systems in which we told the subjects to minimize a cost functional of the form: $\bar{x}^2 + \rho\bar{u}^2$, a linear combination of mean-squared error and mean-squared control movement. We changed ρ and observed how the subject's performance changed. He should have adopted different control strategies for different values of ρ. We observed that the subject changed his control strategy in much the same way that an optimal controller would when ρ changed. However, he did not use the correct values of ρ. In particular, the range of the values of the pilot's subjective ρ was less than the range of the objective ones chosen for the experiment.

DISCUSSION

HAROLD G. MILLER: Your comment on the system and the data reconstruction that represents the actual vehicle hit a nerve, because we have been considering something like this for some time for flight controller's use in the bay wing of the spacecraft.

JEROME I. ELKIND: How will you provide him a data reconstructor if you will?

MILLER: What we do is to circumvent that. Maybe we will have a piece of hardware back at the plant. If we get a problem, we operate it and see how it works.

ELKIND: It is the same idea, really, except in the control model the reconstruction is implemented online, and the data are fed to it.

MILLER: The thing that holds this up is the complexity, cost, and time. It is a pretty big job, really, unless you can prove some positive results out of it to undertake something like this.

ELKIND: Belsley would object if I said "implementing," because we are not implementing. We have just a model. The reconstruction and all the rest are in the pilot. We can use these techniques without worrying about reliability and implementation.

Decisionmaking in Manned Space Flight[1]

R. Mark Patton and Julie A. Rauk

NASA Ames Research Center

The history of the manned space-flight program reveals a trend toward increasingly greater use of the pilot as the primary operator of various subsystems. When one compares the general design philosophy that existed at the beginning of the Mercury program with that of present systems, Apollo, and its proposed successors, the contrast is evident. The original design philosophy of the Mercury program was that all critical functions were to be fully automatic, with man overriding only in the event of system failure. Such things as spacecraft attitude stabilization, retrorocket firing, and drogue and main parachute deployment were made fully automatic. Man was to go along primarily as a passenger and an observer, and the idea of his ever having to "take over" seemed remote to many. I think it fair to say that a decade ago, design engineers typically had more faith in the reliability of fully automatic systems than they do now. If that seems too sweeping a statement, at least it was far easier then than it is today to find someone who would argue that man is unnecessary, because his reliability can never approach that of the machine.

As it turned out, during each of the four manned orbital flights of Mercury, one or more failures in the automatic systems occurred, man was able to take control, and the mission was successfully accomplished. By the end of Mercury, man had proved to be a capable performer in space. Even during the

program, the growing emphasis on manual operation was evident. Prior to MA–9, White and Berry (ref. 1, p. 44) noted that "each flight plan has further reduced the automatic activity and provided more necessary pilot input." Succeeding projects have been designed from the beginning around the concept of greater pilot involvement.

In operating current manned spacecraft systems, the pilot is heavily involved in overall mission or systems management. Compared with aircraft piloting, control of the flight path is a minor part of the total task. By its nature, systems management often requires complex decisionmaking on the part of the operator. When we consider human limitations in spacecraft piloting, these are more likely to be limitations in the information processing/decisionmaking area, rather than limitations in psychomotor abilities as was the case with aircraft. Human error is more likely to occur in decisionmaking than in perceptual-motor activity. My belief is that the potential contributions of decisionmaking research to the manned space-flight program are great. Areas of application include the design of system hardware, the specification of operational procedures, and even the selection and training of astronauts to be effective decisionmakers.

As a starting point, one would like to have a classification scheme covering the decisions that are likely to have to be made by the space vehicle pilot. I have in mind some sort of classification based on the circumstances under which decisions are likely to be re-

---

[1] Paper presented by Mr. Patton.

quired, the nature of the information inputs, and the nature of the required actions. Such an analysis would be valuable in suggesting directions in which our laboratory work ought to proceed. We made what must be considered an abortive attempt to develop a classification scheme by looking at all of the Mercury and Gemini flight reports, attempting to identify decision situations, and further attempting to categorize them in some sensible way. Frankly, we did not get very far. When you try to use them for this kind of analysis you find that the information contained in flight reports is not very detailed, and tends to concentrate on the major and dramatic items. These are the items that get into the newspapers at the time. I am sure that every flight involves a multitude of decisions that are relatively obscure, but no less important and, perhaps, no less complex and difficult than the "big ones."

Because we found few usable examples, I thought that I might describe two outstanding and dramatic examples, simply to illustrate the approach that we were trying. Doubtless you remember the case of the loose heat shield during Colonel Glenn's MA–6 flight (ref. 2). You could list the major characteristics of the situation somewhat as follows:

(1) The inputs were ambiguous. For example, was the heat shield really loose, or was the light that indicated this condition malfunctioning?

(2) The course of action was determined on a probabilistic basis; that is, that the retropack would probably burn off during reentry and thereby cause no difficulty.

(3) Much of the information pertinent to the decision was available only to ground personnel. Thus, there was a need for communication between the ground personnel and the astronaut. As a sidelight, Colonel Glenn felt that the communication that did occur was inadequate. To quote a portion of his flight report (ref. 2, p. 136):

I feel it more advisable in the event of suspected malfunctions, such as the heat shield retropack difficulties, that require extensive discussion among ground personnel, to keep the pilot updated on each bit of information rather than waiting for a final clear-cut recommendation from the ground. This keeps the pilot fully informed if there would happen to be any communication difficulty and it became necessary for him to make all decisions from onboard information.

(4) The required decision and control actions were dichotomous.

(5) A team effort was involved in gathering and evaluating information.

(6) Colonel Glenn's decision posed danger only to himself.

(7) A great deal of time was available before the decision had to be made.

By comparison, consider the decision made by Captain Schirra in an abortive attempt to launch Gemini VI (ref. 3, p. 19). He elected not to eject on the pad when an equipment failure caused cockpit instruments to indicate that liftoff had occurred, but that the missile had generated insufficient thrust to fly, and therefore was settling back onto the stand. Captain Schirra, who had experienced the physical sensations of liftoff during the MA–8 mission, felt that it had not, in fact, occurred in this case. One must remember that the decision to remain with the vehicle had to be made in a very brief time, on the order of a second or so, with catastrophic results if the decision not to abort was wrong. Because ejection did not occur and because the vehicle had not lifted off, it was possible to launch the vehicle only 3 days later, in time for it to rendezvous successfully with the Gemini VIII. In comparison with the MA–6 heat-shield problem, items of similarity were that—

(1) Information was ambiguous.

(2) Probabilistic decisionmaking was involved.

(3) Information was available to ground personnel that was not available to the crew.

(4) The decision and control actions were dichotomous.

Differences were that—

(1) A team effort was impossible; the decision had to be made by one man.

(2) Because there was a two-man crew, Captain Schirra was making a decision about the other astronaut's life.

(3) Unlike the previous case, there was no time to solicit information from ground personnel.

I think that such an analysis, examining both those parameters which may be expected to change from flight to flight and those which may be expected to remain constant, would provide a wealth of material that we could use in pinpointing areas of research need.

I am sure that there are better sources of information upon which such an analysis might be based than the ones we used. One candidate is the failure task analysis. Grether (ref. 4) describes the process:

> One of the more interesting activities of the McDonnell group was their failure task analysis. This consisted of an analysis of possible equipment failures that could occur during flight, as determination of the symptoms which the pilot or ground crew would have of the failure, the appropriate actions to take, and the consequences to be expected.

Such an analysis should provide excellent material for developing the classification scheme that I envision.

We were able to locate a number of articles that deal to some extent with the topic of decisionmaking in space-vehicle operation (see bibliography). Most simply contain passing references, stating such opinions as that decisionmaking is an important aspect of astronaut activity, and that man's decision capability is the primary reason for his being sent into space. Few authors attempted to be very specific on the subject.

One document (ref. 5) dealt extensively with the maintenance of alertness during manned space missions and the development of means to monitor the crewmembers' state of alertness. The role of alertness in influencing the quality of decisionmaking is emphasized. The author suggests that objective indicators can be developed that would be calibrated to each crewmember. One or more "key" indicators, or some weighted combination of indicators, would be developed using data gathered in simulation exercises performed under both normal and stressful (such as fatigue) conditions. This is, of course, a statement of a particular approach to the universal question of objective performance measurement in real-life systems operation, and is always easier to propose than it is to do. However, it seems to me that the author has some good ideas regarding possible approaches to monitoring techniques, and has kept his suggestions within the bounds of what seems possible in the manned space-flight programs.

Thus far I have spoken only of complex decisionmaking. The range of behavior that can be thought of as having at least an element of decisionmaking involved is very broad indeed, as we have noted several times during this conference. Jerison and Pickett (ref. 6) attempt to account for performance on vigilance-type tasks (that is, signal detection tasks where signals are weak, infrequent, and temporal uncertainty is great) in terms of the decision processes involved. This article attracted my attention, and I report it here because the authors attempt to relate their research and theories to decisionmaking behavior in manned space flight.

Tasks requiring vigilance are well represented in spacecraft. In Jerison and Pickett's words (ref. 6, p. 211):

> Vigilance remains a human-factors problem in space missions because there may be no alternative to man as a monitor. This might occur accidentally if there is a breakdown in an automatic detection system, or unintentionally if the design delays or weight penalties for using automatic detection equipment are great and the penalty for an occasional missed signal is moderate. Furthermore, the man in a space vehicle will want information on the status of the vehicle's systems, even if these systems are intended to be automatic and self-correcting. Such supplementary information systems might most conveniently be visual diplays that are monitored only occasionally.

Jerison and Pickett take a particular view of the vigilance problem: that the quality of performance is determined primarily by the observer's decisions on whether or not to attend to the display. They feel that this contrasts, for example, with a notion that there are changes in the observer's criterion of signal/no-signal during the vigil. Their particular interest is in cases in which the probability of a signal's appearing is so low

as to be virtually nonexistent. They believe that although the penalty for a missed signal would be high, perhaps death, the "essentially zero probability for signal appearance . . . would result in a zero expected value for observing." Their guess is that if an astronaut were assigned to monitor a display, with the "best bet" being that no signal would ever appear, "then he and his colleagues would redesign their jobs to eliminate the monitoring task." Perhaps their response would not be so extreme, but I do not doubt that the frequency of monitoring could become very low, thus the probability of a missed signal high, whether the decision involved were explicit or, as seems more likely, implicit. Jerison and Pickett make several suggestions of ways in which procedures might be altered to accord with what we know of decision and vigilance behavior as a result of laboratory investigations and still satisfy operational requirements. These involve either presenting artificial signals to increase signal probability, increasing system sensitivity, or, where possible, storing signals on tape for occasional review in "fast time."

I want to mention briefly selection and training as these relate to decisionmaking. From the beginning, people involved with the selection and training of astronauts have been concerned with man's decisionmaking ability. In a review (ref. 7, p. 173) of factors considered in the selection of the original seven astronauts, Voas cited as a criterion, "a good ability to make decisions," but he did not elaborate on the method used to measure this ability. Perhaps the fact that this was listed as a "personality factor," rather than as one of the "aptitude and ability factors," suggests that it was not based on any specific test, but was more a matter of general impression. There is a widespread belief that engineering test pilots, especially those still living, are particularly adept at making correct decisions in complex and rapidly changing situations. Because all of the early applicants, and most of those coming later, were of this profession, then the problem would be one of making a choice among individuals coming from such a well-

qualified population. The development of better methods of accomplishing this could be a useful contribution to the space program.

The failure-task-analysis approach has been used both in the ongoing process of design refinement, and in the development of crew-training procedures. On the basis of this analysis, and subsequent simulation of the various potential failures, attempts have been made to take as many potential emergencies as possible out of the high-level decision category, by reducing perceived ambiguity and uncertainty and by training the astronauts so that response to foreseen emergencies becomes virtually automatic. In Yntema's terms, attempts are made to move as many potential situations from the "explicit weighing and balancing" type to the "pattern recognition" type. Simultaneous failure of components, partial failures leading to ambiguous cues to the malfunction, and possibly the occurrence of the totally unexpected, suggest that explicit weighing and balancing must remain a factor to be considered. Voas (ref. 8, p. 114), asserting a limited value of ground simulation in maintaining high-level decisionmaking skills, noted that in ground simulation "the penalty for failure is merely the requirement to repeat the exercise," and doubted "whether skill in making such decisions can be maintained under radically altered motivational conditions." This reasoning led to the inclusion of high-performance aircraft piloting in the astronaut training program, on the grounds that skill in high-level decisionmaking would best be maintained in a situation in which danger is real and motivation is high. This presumes a great deal of transfer from one situation to the other.

Occasionally it has been suggested that formal, general training in decisionmaking skills be included in the training program. I do not know if a suitable course exists at this time, but I believe that one could be developed. I do not think that we know enough at this time to state with assurance that such a course would improve on the present, less formal methods.

I will close with a few thoughts concerning

the future. It seems to me that for a long time to come many NASA manned space missions (extended Earth-orbiting missions and any exploration of the near side of the Moon) will be accomplished using techniques, particularly mission-control operational procedures, substantially like those in use at the present time. At least early expeditions to the far side of the Moon, and any interplanetary voyages, will require that the present approach to mission control be altered. Transmission delays will be very long, and the rapid exchange of information between Mission Control Center and the vehicle that occurs on present missions, particularly when emergencies arise and complex decisions must be made, will not be possible.
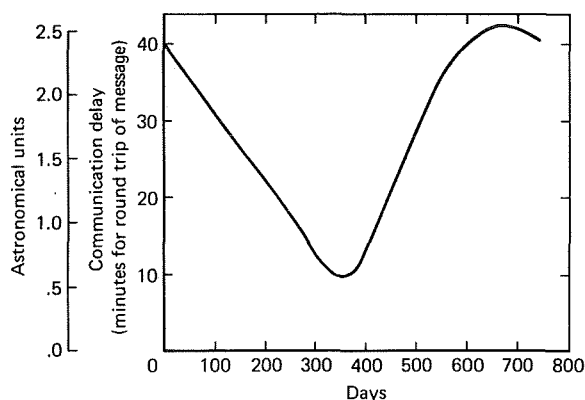

FIGURE 2.—Mars-Earth distances.

Earth of almost 2 astronomical units at a maximum, which represents a round-trip communication of approximately 2000 seconds. The shorter duration mission involves less distance, only slightly over 0.8 of an astronomical unit, but is less attractive from the standpoint of power requirements. Figure 2 plots the distance of Earth and Mars from each other, varying in a cycle whose period is almost 2 years in duration. These data would be applicable to communications between a permanently manned Mars station and the Earth. With such delays inevitable, many more decisions will have to be made by the vehicle crew, solely on the basis of onboard information, than is the case at present. Team decisionmaking is now, and will continue to be, an area of great importance in research. Also, techniques might be developed that would permit the early identification of impending decision situations, perhaps on a probabilistic basis. Then, prior to its actually being required in the decision process, information could be requested by the spacecraft crew, and transmitted back to them, when both line-of-sight communications and adequate time are available.
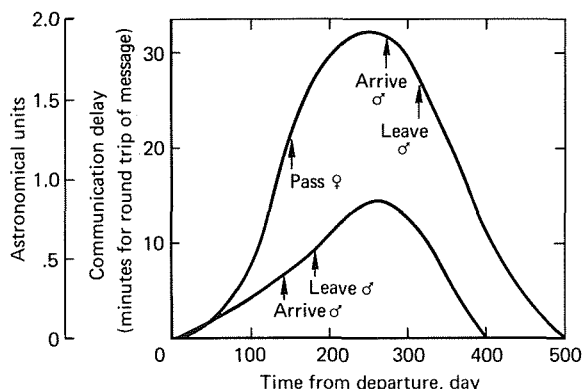
FIGURE 1.—Spacecraft-Earth distances during Mars missions.

With the help of William Allen of the Mission Analysis Division at Ames Research Center, we have prepared two figures that show the duration of the communication lags involved in typical Mars missions. Figure 1 shows two such missions, plotting the round-trip communications lag at various times in the mission. The longer duration mission involves distances of the spacecraft from the

REFERENCES

1. WHITE, STANLEY C.; AND BERRY, CHARLES A.: Résumé of Present Knowledge of Man's Ability to Meet the Space Environment. Aerospace Med., vol. 35, Jan. 1964, pp. 43–48.

2. GLENN, JOHN H., JR.: Pilot's Flight Report. In: Results of the First United States Manned Orbital Space Flight (NASA, Manned Spacecraft Center, Houston) Feb. 20, 1962, pp. 119–136.

3. ALEXANDER, GEORGE: Fast Analysis Overruled Gemini 6 Ejection. Aviation Week and Space Technology, vol. 82, no. 25, 1965, pp. 19–20.

4. GRETHER, WALTER F.: Psychology and the Space Frontier. American Psychologist, vol. 17, 1962, pp. 92–101.

5. ANON.: Alertness in Manned Space Flight: Part II: Operational Guidelines. NASA CR–85496, Jan. 29, 1964.

6. JERISON, HARRY J., AND PICKETT, RONALD M.: Vigilance: A Review and Reevaluation. Human Factors, vol. 5, 1963, pp. 211–238.

7. VOAS, ROBERT B., ET AL.: Astronaut Training. NASA SP–45, 1963.

8. VOAS, ROBERT B.: Project Mercury: Astronaut Training Program. In: Psychophysiological Aspects of Space Flight (Bernard E. Flaherty, ed.), Columbia University Press (New York), 1961, pp. 96–116.

BIBLIOGRAPHY

BOYNTON, JOHN H.: The Pilot's Role During Mercury Systems Failures. [Preprint 64–222], paper presented at First AIAA Annual Meeting (Washington, D.C.), June 29–July 2, 1964.

CHRISTENSEN, JULIAN M.: Performance Capabilities of Man on Earth and in Space: A Study of Continuity. Advan. Astronaut. Sci., vol. 20, part I, 1965, pp. 37–56.

DEAN, ROBERT: Human Stress in Space. Sci. J., vol. 153, no. 3743, 1966, pp. 71–75.

HARTMAN, BRYCE O.: Time and Load Factors in Astronaut Proficiency. In: Psychophysiological Aspects of Space Flight (Bernard E. Flaherty, ed.), Columbia University Press (New York), 1961, p. 278.

HODGES, JOHN D., JR.: The Decision-Making Function in System Simulation—An Approach. In: National Symposium on Human Factors in Electronics, 5th IEEE, 1964, pp. 294–303.

KRAFT, CHRISTOPHER C., JR.: A Review of Knowledge Acquired From the First Manned Satellite Program. In: Virginia Polytechnical Institute Conference on Artificial Satellites, pt. C, 1964.

MATHEWS, CHARLES W.: United States Experience on the Utilization of Man's Capabilities in a Space Environment. In: Life Sciences and Space Research (R. B. Livingston, ed.), John Wiley & Sons, Inc. (New York), 1963, pp. 141–159.

McDIVITT, JAMES A.: Rendezvous, Rescue and Recovery Aspects of Space Operation. In: Advances in the Astronautical Sciences, vol. 16, pt. II, 1963, pp. 12–19.

RATH, GUSTAVE J.: Information Overload During Reentry. Proceedings, Annual Meeting of the Institute of Environmental Sciences, 1962, pp. 349–352.

RITLAND, O. J.: Need for Man in Space. In: Lectures in Aerospace Medicine, pp. 61–76 (USAF School of Aerospace Medicine, Brooks Air Force Base, Texas), Feb. 4–8, 1963.

VOAS, ROBERT B.: A Description of the Astronaut's Task in Project Mercury. Human Factors, vol. 3, Sept. 1961, pp. 149–165.

Models for Memory[1]

RICHARD C. ATKINSON AND RICHARD M. SHIFFRIN

Institute for Mathematical Studies in the Social Sciences
Stanford University

This paper includes a fairly general theoretical framework for human memory, beginning with an outline of the theoretical system and followed by some specific experiments and models derived from the overall framework. A great deal of research has been carried out in recent years on human memory, especially short-term memory, and the models proposed by various researchers have begun to dovetail into a commonly accepted theory.

The most important theoretical distinction of the system is that between structural features of memory and control processes. The structural features are permanent and include both the physical system and built-in processes that do not vary from one situation to another; for example, the hypothesized short- and long-term memory stores. Control processes, on the other hand, are selected, programed, and used, at the option of the subject. The use of a particular control process at some moment will depend on such factors as the task, the instructions, and the subject's particular response history. Examples of control processes are coding, mnemonics, visual imagery, and rehearsal strategies.

Control processes were examined extensively around the turn of the century, primarily because they are what the subject reports when asked to describe what he is doing in a particular task. For many reasons, consideration of control processes tended to drop into disfavor, and even today some experimenters hesitate to ask their subjects to introspect about what they are trying to do. It is often assumed that the various strategies used by the subjects are not important, or that they vary enough from subject to subject to randomize away any consistent effects in large groups. The point I should like to emphasize is that this position is wrong: control processes are in most cases extremely important determiners of performance in human-memory experiments. Different control processes not only affect the level of performance but also the functional relationships found. As an example, consider a series of experiments examining mental imagery and coding by Bower (ref. 1) at Stanford University. In the course of these experiments, subjects are told to encode verbal material by forming vivid mental images. Compared with control subjects, the subjects' performance was higher by a factor of 5 or more. Their rate and form of forgetting also were markedly affected. Furthermore, these subjects continue to use these techniques in future experiments they enter. As a result, it has been necessary to ask prospective subjects for our experiments whether they have previously taken part in Bower's experiments and, if so, eliminate them. This is one example of the importance of control processes. Experiments will be described in which per-

formance is primarily dependent upon a subject-controlled rehearsal strategy called the "buffer."

Having distinguished between processes and structural features, I now want to categorize the memory structure into three components: the sensory register, the short-term store, and the long-term store (see fig. 1). It is possible to further subdivide these components on the basis of the sensory modality of the stored information, for there is clear evidence indicating that the characteristics of memory for different sensory inputs may differ considerably. In this paper, however, the distinction will not be emphasized, for it complicates the presentation.
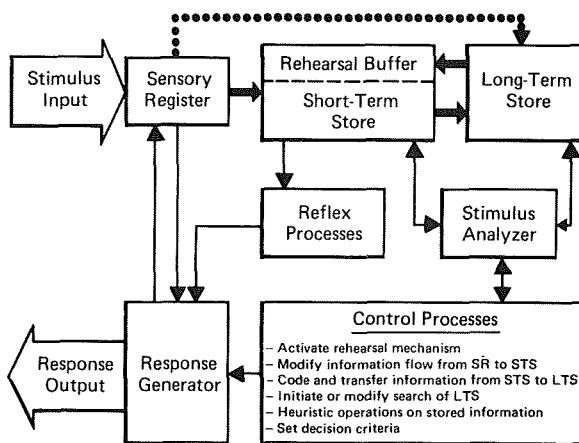

FIGURE 1.—The theoretical memory system: structural distinction between memory and control.

The sensory register accepts incoming sensory information and holds it fairly accurately for a very brief period of time, perhaps several hundred milliseconds; the information then decays and is lost. The short-term store is the subject's working memory; it receives selected inputs from the sensory register and also from the long-term store. Information in this store decays and is lost in a period of about 30 seconds or less, but may be maintained via control processes, such as rehearsal, as long as the subject desires. The long-term store is a fairly permanent repository for information, which is transferred from the short-term store. Note that trans-

fer does not imply that information is removed from one store and placed in the next; rather transfer is used here to imply the copying of selected information from one store into the next without affecting it in the original store. Note, also, that the term "information" is used in a nontechnical sense; it does not refer to bits but to the various sounds, codes, and so forth, that are stored.

We now turn to an outline of the characteristics of the three memory stores, and a description of the major control processes associated with each.

The prime example of a sensory register is the visual image investigated by Sperling, Averbach and Coriell, Estes and Taylor (refs. 2, 3, and 4, respectively), and others. If an array of letters is presented on a T-scope and the subject is asked to report as many letters as possible, usually about six letters are reported; even a 30-second delay until report does not affect performance. It appears that the subject is transferring about six letters from his sensory register to an auditory short-term store where they are rehearsed until a report is requested. This hypothesis is in line with introspective reports and the fact that response confusions tend to be auditory rather than visual. In order to examine the visual image itself, partial-report procedures have been used. For example, a matrix of letters is presented and, a brief period after the tachistoscopic display, a tone signals the subject as to which row of letters to report. When the delay until the tone is very short, performance is virtually perfect. As delay increases, performance decreases, reaching asymptote after several hundred milliseconds. In addition to this decay with time, it has been found that succeeding visual stimulation modifies preceding images. At present, not much is known about the form of the decay; whether the letters decay together or individually, probabilistically or temporally, all or none or continuously. Transfer to the short-term store takes place during the period before the image decays: the limited amount transferred is not so much dependent on the time needed to scan the letters (about 10 msec) as the capacity

of the short-term store itself. The dotted line in figure 1 indicates a noncommittal attitude as to whether direct transfer to long-term store takes place from the sensory register. At least it is clear that there exists direct communication between the sensory register and the long-term store, because a visual letter may be transferred to an auditory short-term store. In order for this to occur, the auditory representation of the visual image must be retrieved from the long-term store and then placed in the short-term store.

Control processes relevant to the sensory register include the following: the decision as to which sensory register to attend to if several are activated simultaneously; where and what to scan within the decaying image; how to perform the long-term matching of the information residing in the sensory register, if a search is being made of the decaying image; and what information to transfer to the short-term store.

The short-term store is the next feature to be considered. Discussion will be restricted to the verbal-linguistic short-term system rather than, say, visual or kinesthetic short-term memories. The short-term store is viewed as the subject's working memory because control processes are based within it and directed from it. For a number of reasons, the characteristics of the short-term memory trace are difficult to determine. For one thing, rehearsal processes can maintain selected information for indefinitely long periods. For another, while an item resides in the short-term store, information about it begins to accumulate in long-term store, and consequently test performance will be a joint function of retrieval from both stores. Some evidence has been accumulated however; Norman and Wickelgren (ref. 5) have carefully controlled rehearsal in a task where subjects listen to rapid sequences of three-digit numbers. Following the sequence, a three-digit number is presented for test and the subject must decide whether the test item was in the presented list or not. These data were accurately fit by assuming an item reached a given strength level, depending upon the

presentation time, and that this strength then declined exponentially with the number of succeeding items. In this situation, the number of items rather than time per se was the important variable determining decay. Other experiments, such as Peterson and Peterson (ref. 6), in which a single-consonant trigram is followed by arithmetic and then attempted recall, have indicated that short-term decay takes place in about 30 seconds or less whenever rehearsal is inhibited.

Of the many control processes in the short-term store, I shall be primarily concerned with rehearsal mechanisms. Henceforth, "rehearsal" is taken to refer to the maintenance of information in short-term store, through vocal or subvocal repetition. That is, each decaying trace in short-term store is, so to speak, reset by each repetition, whence it begins to decay again. A little consideration makes it clear that the optimal rehearsal method for maintaining a maximum number of items in short-term store occurs when a fixed set of items is being rehearsed at any one time, probably cyclically, such that each item is rehearsed just before it decays. This kind of fixed-size rehearsal set is called a buffer and is shown in figure 2. The buffer consists of a selected portion of the information currently in the short-term store. At any one time, there may be information in the short-term store not in the buffer, but this information is rapidly decaying, whereas that in the buffer is maintained by rehearsal as long as the subject desires. The size of the buffer, or the number of items concurrecently rehearsed, is denoted by r; r will, of course, depend on such factors as the nature of the items being rehearsed; if each item is an eight-digit number, r might be 1, whereas if each item is a single digit, r might be 8. For verbal rehearsal, it seems likely that r is absolutely determined by the auditory length of the rehearsal material. This hypothesis is consistent with the estimates of r recovered from our experiment. In most experiments with a continuous sequence of presented items, the subject must decide when an item is presented whether to enter it into the
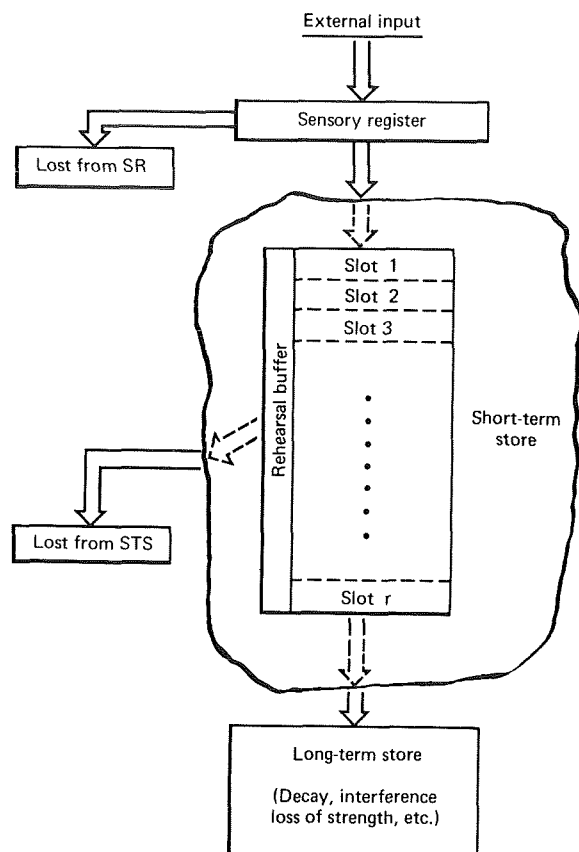
FIGURE 2.—Fixed-size rehearsal set: a buffer.

buffer or not; if it is entered, then the subject must decide which item currently undergoing rehearsal should be removed to make room for the new entering item.

It should be emphasized that this rehearsal buffer is a control process set up at the will of the subject in appropriate experimental situations. Factors such as difficult items, short study-test intervals, and complex or rapid presentation rates will tend to induce subjects to utilize a buffer and will determine its size, r. On the other hand, an experiment in which word-word paired associates are to be learned over an extended number of trials will probably result in the subject coding rather than rehearsing.

The short-term system I have been describing is not as easy to examine experimentally as one would like. One problem is that during an item's stay in the short-term store, infor-

mation about it accumulates in the long-term store; thus the probability of a correct response will always be a combination of the retrieval probabilities from both stores. The working of the short-term system would be much easier to examine if long-term storage could somehow be blocked. For this reason it is interesting to consider some remarkable results reported by Milner (ref. 7) who accomplishes this effect surgically. She reports that patients given large bilateral hippocampal lesions demonstrate a marked memory deficit, being unable to store new information permanently, or to recover such information if it is stored. Thus a patient, if he changed addresses following the operation, will, if left alone, return to the old address, not knowing he has moved. This inability to store new information is quite general, the only possible exception being an ability to improve in some simple rote motor skills. On the other hand, information stored preoperatively is recovered without difficulty. Furthermore, the patients have an unimpaired short-term store so that they demonstrate unchanged IQ as measured by standard tests. If given a sequence of instructions to carry out, they can do so by rehearsing the instructions, but immediately forget if rehearsal is interrupted. If given various material to remember for a short time period, they can do so as long as the material is verbally encoded for rehearsal. If interrupted, however, the material is forgotten. If the material presented is not easily encodable (such as tones and figures), then memory of this material is lost in about 30 seconds, even if there is no interruption. Incidentally, these operations, originally performed to eliminate seizures, have been discontinued. Nevertheless, the effects have been validated in other ways. For example, sodium amytal injected into the carotid artery will temporarily knock out one hemisphere—if the other hemisphere has a hippocampal lesion, or damage, then the patient exhibits the above memory deficit. Furthermore, patients suffering from Korsakoff's syndrome tend to exhibit the same effects and this syndrome is often associated with hippocampal damage.

All in all, these results tend to lend considerable support to the memory system previously outlined, particularly in regard to the division into short- and long-term memory stores, and also the characteristics, such as rehearsal, of the short-term store.

Next consider the transfer of information between the short- and long-term stores. Quite often transfer of information occurs from the long-term to the short-term store, as during coding, hypothesis testing, stimulus naming, and so forth. However, the transfer from short- to long-term store is of greater interest at the moment. The assumption is made that at least some transfer takes place during the period that any information resides in the short-term store, regardless of whether any active attempt is made to store or not. Experiments supporting this assumption were performed by both Hebb and Melton (refs. 8 and 9, respectively): subjects were given a series of digit span tests, in each of which they had to repeat a series of digits read to them. There was no reason to try and learn the individual sequences. Unknown to the subjects, however, some of the digit sequences were repeated at intervals. Without the subjects noticing the fact, performance on the repeated sequences improved over trials. Apparently a long-term trace was being stored and strengthened for the individual sequences, as a result of brief rehearsal in the short-term store. Although we assume that storage always takes place during an item's residence in the short-term store, the amount and the form of the transferred information are markedly affected by the control processes used. If the subject is using simple rehearsal, the long-term trace stored may be only a weak auditory image of the repeated item. Coding, on the other hand, might result in the storage of a rich visual image, containing not only the information in the stimulus input but much more besides.

The final major structural component of the system is the long-term store. Although it is possible to subdivide this store on the basis of physiological evidence pertaining to consolidation, for present purposes we will

view it as a single store, obeying fixed rules throughout. The two basic questions to be answered are: "What is stored?" and "What happens to the memory trace as time passes and other information is stored?" The form of the memory trace will be dependent on the control processes governing the transfer from the short-term store. Most generally, the memory trace may be viewed as a multicomponent information array, stored in either one place or in many (single versus multiple-copy model). Somewhat less general, but more useful in applications, is the representation of the trace by a numerical strength value, the higher strengths indicating storage of greater amounts of available information. Restrictions of these general models for the form of storage lead to specific alternatives that may be appropriate in individual situations, depending both upon the nature of the task and also the control processes invoked by the subject for transfer, search, and retrieval. An example is the simple all-or-none one-element model, in which a single complete information copy is either in memory or not, and once stored, stays stored. This model has been used appropriately in fixed-list, paired-associate experiments in which the responses contain minimal information, often consisting of the two letters, a and b. Other models that have been applied include single-memory copies of partial information; many-memory copies, each of complete information; and many-memory copies, each of partial information. It should be emphasized that the question is not which of these versions is correct; rather it should be asked which model is appropriate for a given situation, the answer depending upon the nature of the task and the type of control processes that the subject is induced to use.

The next feature to be considered is the change in the memory trace in long-term storage and its retrievability over time and intervening material. In considering this topic, we are plunged into the large body of literature roughly classified by the term "interference theory." In the present discussion, however, I will use the term "interference"

in a restricted manner to refer to the decay or destruction of information over time. Thus, interference models will assume permanent loss of information occurring between study and test, and forgetting can be viewed as a structural feature of the memory system. At the other extreme are the models assuming no loss of information over time; rather, forgetting occurs because of the decreasing effectiveness of the memory search at the moment of test. Note that the usual theories of forgetting, such as decay theory, response competition, unlearning, and so forth, can be liberally interpreted so as to be consistent with either of these models. Unfortunately, it is not easy to distinguish between the two models: structural change occurring during delay, on the one hand, versus decreased search effectiveness as a result of increased information storage. Each has a certain amount of face validity. Gross physical considerations of nerve-cell destruction, seizures, conditions of anoxia, and so forth, make it virtually certain that at least a small amount of information is lost over time. On the other hand, such effects might be relatively negligible in the typical memory experiment. Certainly the assumption that stored information is never changed or lost has a certain elegance. Furthermore, long-term memory search is one of the most common subjective features in everyday experience. Slightly less obvious than the fact that memory search exists is the fact that the search routine is very highly under the control of the subject, in terms of where in memory to search, how to search, what criteria to set for terminating the search, and so forth. If one is asked to name the countries of Asia, for example, one can search memory in a free associative fashion, or alphabetically by first letter, or geographically, to name just a few mechanisms. In general, the various search mechanisms will result in markedly different performance. In a test-retest procedure, for example, a random search for the names of the countries of Asia will result in differing responses, whereas an ordered geographic search will result in essentially the same output in the same order.

Our discussion of long-term storage began with the consideration of the structure of long-term forgetting. In reference to the large body of literature, we will simply state here that the pattern of results is highly complex and not amenable to simple description. This state of affairs is likely to remain until a better grasp is gained on the control processes involved, such as what the subject thinks he is to remember, what he rehearses, how he codes, and how he searches memory at test.

This concludes a very brief outline of the memory system. Although much of the system is quite tentative, and some areas have barely been touched, the main gist of the theory should be relatively clear. I will next describe some specific experiments and models that should illustrate applications of the overall system, and also how certain types of control processes can be dealt with quantitatively.

The first experiment is easy to describe. A series of cards was placed in a row before the subject. Each card had a colored patch on one side, the color always picked randomly from four possibilities. The cards were presented at a rate of one every 2 seconds; as each card was presented, the subject called out its color and the card was then turned over. The list length was varied from trial to trial, but the subject always knew prior to the start of a list how many cards would appear in that list. At the end of the display, the experimenter pointed to one of the cards and the subject guessed the color under it. He also gave a confidence rating from 1 to 4 of the probability that his guess was correct, 1 being most confident.

The display size took on the values 3, 4, 5, 6, 7, 8, 11, and 14. Figure 3 shows the probability correct as a function of serial position for the different display sizes. The most recently presented item is to the far left. The circles are the observed data. Displays of sizes 3 and 4 were always responded to correctly and are not graphed. The major factors to be explained are the strong recency effect to the left of each figure, with its tendency toward an S-shape; the less-strong
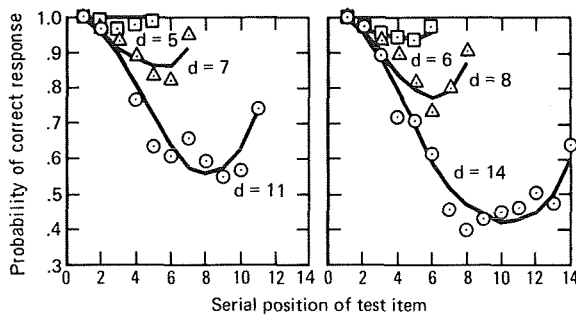
FIGURE 3.—The probability of a correct response as a function of serial position for the different display sizes.

primacy effect to the right; and finally, the decrease in performance as list length increases.

Now, what kind of model is applicable here? Because each color is read aloud when presented, we assume every item successfully passes through the sensory register and enters the short-term store. Within the short-term store, there are a number of a priori reasons for expecting the subject to use rehearsal as his primary strategy. For one thing, long-term coding would be inefficient because something well-learned for one display would undoubtedly confuse the subject on immediately following displays. For another, the very short displays can be handled perfectly by rehearsal alone, and this strategy may then be carried over to longer displays. Finally, a posteriori, subjects report rehearsal as their usual method of operation. For these reasons, we shall apply a version of the buffer model described earlier. That is, it is assumed that a fixed number of items is being rehearsed at any one time. For any presented item the subject must decide whether to rehearse it or not. In other studies where subjects used a rehearsal scheme, it was found that the act of reading each item aloud as it was presented led them to enter every item into rehearsal. Because the colors were read aloud by the subject in the present case, we shall assume that every item enters the rehearsal buffer. The initially presented items fill the buffer; once the buffer is filled to capacity, the next

entering item must take the place of an item already in the buffer. It could be assumed that the item to be lost is chosen randomly; however, this model will predict an exponentially decreasing recency effect rather than the S-shaped one found. The recency effect arises, of course, because any item in the buffer at the time of test is reported correctly. The observed S-shaped effect can be predicted if it is assumed that there is a tendency for the items that have been undergoing rehearsal for the longest time to be the ones lost from the buffer first. This is actually a fairly rational strategy for the subject to use; for example, the longer the items remain in the buffer, the more strength is built up in long-term store for them. Under fairly general conditions, it is an optimal strategy to drop from rehearsal the item that has accumulated the greatest amount of longterm strength. In order to quantify the notion of a tendency for the oldest item to be the first lost, we let P_i be the probability that the ith oldest item is the one dropped from rehearsal, and we set

$$P_i = \frac{\delta(1-\delta)^{i-1}}{1-(1-\delta)^r}$$

where δ is a parameter to be estimated and r is the buffer size. Note that this function has appropriate properties: if δ approaches 0, then a random item is dropped from the buffer; if δ is 1, then the oldest item is always dropped from the buffer.

The model as described so far, without any long-term storage, will predict an S-shaped decreasing recency effect for each list length. It will not predict a primary effect and it will not portray the decrease in performance with increasing list length.

The transfer of information to the long-term store will be assumed to take place as a linear function of the total time an item undergoes rehearsal. This is the typical assumption we use in experiments in which the subject utilizes a rehearsal buffer, but would not necessarily apply if coding or some other strategy were used. In particular, transfer of information takes place at a rate θ per unit of rehearsal time, where θ is a

parameter to be estimated. The retrieval assumptions are that, at test, the subject first searches the buffer. If the item is in the buffer, it is reported correctly; if the item is not in the buffer, a search is made of long-term store, and the probability of recovery will depend on the amount of information stored. In particular, recovery will be an exponential function of the amount of information stored. That is, $P(R) = 1 - \exp(-I)$, where I denotes the amount of stored information. This function has the appropriate properties: when I is 0 the probability of recovery is 0, and if $I \to \infty$ then $P(R) \to 1$. If retrieval is not successful, then the subject guesses. The model to this point will now be able to predict the primacy effect in the data. More information is stored about early items in the list, both because they reside longer in the buffer (while the buffer is filling) and because they receive more rehearsal time before the buffer is filled. (If n items are in the buffer, each is rehearsed $1/n$th of the time.) This model still fails, however, to predict the decrease in performance as list length increases. The reason is evident, because no consideration of interference effects, or of long-term search problems, has been undertaken. Clearly, the probability of recovery from long term should decrease as the number of items increase; either because interference from other items will reduce the amount of information stored, or because a search through an increasing total amount of stored information will be increasingly less effective. Of these two possibilities, we choose to quantify the first. The simplest interference assumption possible is made: for any item in the list, each item preceding it, and each item following it will cause its stored information to be reduced by a proportion τ, where τ is a parameter to be estimated. The model is now ready to be applied. It has four parameters: r, the buffer size; δ, the tendency for the oldest item to be dropped first; θ, the information transfer rate; and τ, the proportional loss of information due to interference. A minimum χ^2 technique was used to pick the best set of parameters:

$r = 5$, $\delta = 0.38$, $\theta = 2.0$, and $\tau = 0.85$. The predictions are the solid lines in figure 3. That the predictions are accurate is indicated by a χ^2 of 44.3 on 42 degrees of freedom.

If this model is really accurate for this situation, we should expect that it could be applied to the confidence-rating data also. Consider confidence rating 1 (most confident), for example. The natural and simplest assumption would hold that any item in the buffer is always given a confidence rating of 1. The probability of giving a confidence rating of 1 for an item recovered from long term would be a function of its strength similar to that for probability correct; that is, the more strength, the higher the average confidence rating. If these assumptions are correct, serial position curves for confidence ratings of 1 should be very similar to the probability correct curves. Figure 4
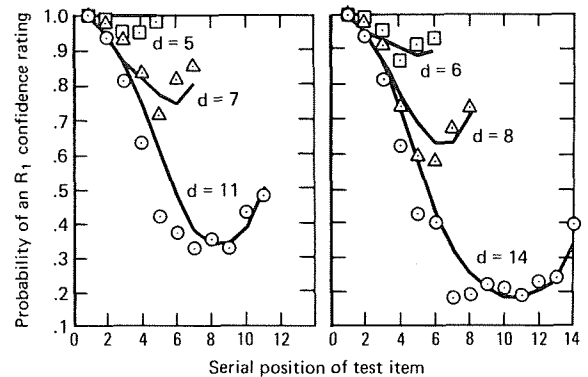
FIGURE 4.—The probability of an R_1 confidence rating as a function of serial position of the test item.

shows the data. One new parameter was used to generate predictions: this parameter determining the probability of giving a confidence rating of 1 for a given amount of long-term strength. The predictions are fairly accurate, as can be seen. Not graphed are the curves for the other confidence ratings, but, as predicted (because they do not include items from the buffer), they demonstrate a missing recency effect. It is impossible in this paper to consider all the various alternative models that might be applied to these data, but one example

is rather interesting. Suppose a model is proposed with only one memory store, containing traces that rise and fall in strength according to unspecified rules, but rules that somehow or other generate the probability correct curves. Such a model would predict that items at different serial positions that have the same probability correct should have the same distribution of confidence ratings, or at least distributions similar in structure. Such is not the case, however. For example, positions 6 and 14, for 14-item displays, have about the same probability correct; as seen in figure 4, they also have about the same proportion of confidence ratings of 1—about 0.40. Position 6, however, has a proportion of confidence ratings of 2 which is 0.36, a decrease from 0.40; position 14 has a proportion of confidence ratings of 2 which is 0.48, a large increase. This is, of course, expected from the buffer model, because position 6 has a fair proportion of recoveries from the buffer, whereas position 14 has almost none, but the result is not easy to reconcile with a single-memory assumption.

Of the assumptions of this model, that one most susceptible to alternative formulation is undoubtedly the one dealing with the decrease in stored information in long-term store. This assumption stated that all other items in the list caused a proportional decrease in stored information. However, this interference postulate may readily be replaced by one in which a subject-controlled search through an increasingly large store of information becomes progressively more difficult. In order to gain a perspective on this point, we turn to an entirely different set of experiments, those referred to as free-verbal recall.

In a free-verbal recall experiment, a list of words is presented to the subject one word at a time at a fixed rate. Following this presentation, the subject recalls as many words from the list as he can, in any order. Figure 5 shows some typical results from Murdock (ref. 10). Graphed is the probability correct as a function of serial position. Note that the most recently presented
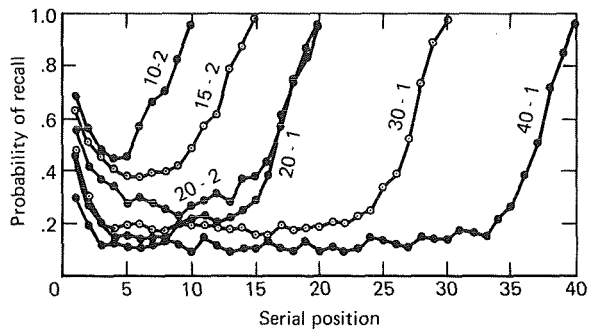
FIGURE 5.—The probability correct as a function of serial position.

item is plotted to the far right, just the opposite of the plots in the previous study. The list lengths are 10, 15, 20, 30, and 40, and the presentation rates are either one word per second or one word every 2 seconds. It is easy to see in these curves a marked similarity to those from the previous study, and a natural hypothesis would hold that a similar model should apply. Is it reasonable in this case that subjects utilize a rehearsal buffer? Several factors make it seem likely: First, rehearsal, if used, will contribute a large proportion of the performance actually observed, because subjects only report about eight words on the average, and the buffer can hold 4 or 5; the rehearsal will be efficient. Second, the fast presentation rates used will make coding relatively difficult. Third, many subjects report using rehearsal mechanisms of the sort hypothesized. Finally, one can use experimental manipulations to examine effects of the short-term store in this situation. Remember that the recency effect (to the right on figure 5) arises because items in the buffer at test are reported correctly. If, following presentation, a manipulation is performed that will cause items in the buffer and short-term store to be lost, then the observed data should no longer have a recency effect; in particular, the primacy effect and the asymptote (in the central portion of the curves) should remain untouched, but the recency effect should disappear. Experimentally, the short-term store has been

emptied by having the subject perform 30 seconds of arithmetic following presentation of the list; then recall is requested. Figure 6 shows results gathered by Postman and Phillips (ref. 11) using this technique. Figure 6(a) shows the standard results without arithmetic; (b) shows the results with 30 seconds of intervening arithmetic. List lengths of 10, 20, and 30 are shown. As predicted, the recency effect disappears but the rest of the curve remains unchanged. It therefore seems quite reasonable to apply a buffer model to this situation.

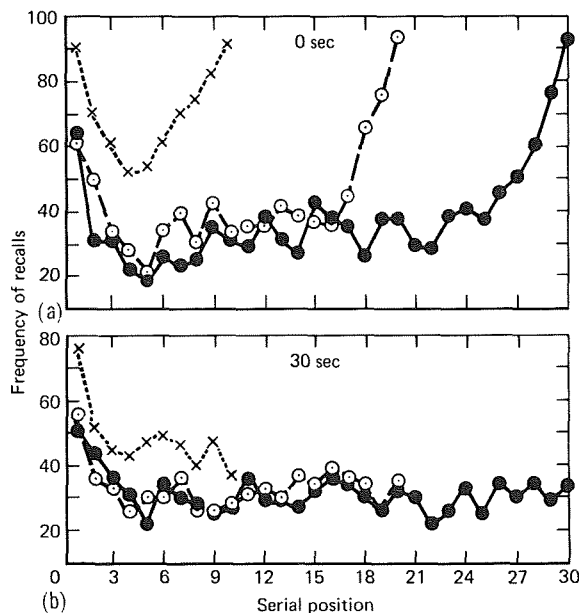FIGURE 6.—Frequency of recalls as a function of serial position. (a) Standard results without arithmetic; (b) results with 30 seconds of intervening arithmetic tasks.

The precise model to be used is similar to that used in the previous study but even simpler. The buffer size is r. Every item is assumed to have entered the buffer. Items lost are chosen randomly, rather than with a biased probability as in the previous study. The reason for the change is primarily that we are not going to try to fit the recency portions of the curves (although it could be done) in the present analysis. Long-term information is again assumed to build up

as a function of time spent in the buffer. Changes in presentation rate are thus reflected in amounts of long-term information built up. In this model, however, interference is not assumed to occur; rather a search process is invoked in order to account for the observed decrease in performance as list length increases. This model was originally proposed because it seemed quite obvious that the task required an extended search of long-term memory on the part of the subject. The strong assumption is thus made that no information stored is lost.

The search scheme is assumed to be, to a large degree, random. Suppose that, at test, the strengths in memory for different items are S_i and the total strength is $\Sigma S_i = S$. We assume that in his search the subject makes n independent picks into the total stored pool of information, where n is determined primarily by the time allotted for responding. On any one pick, the probability of finding the information relevant to item i is simply S_i/S. Having found this information, the probability of recovering the correct word will depend upon the strength S_i. As usual, the recovery probability will be set equal to an exponential function of the strength. In this model, the presentation rate affects the strength S_i for an item and, hence, affects the recovery probability following a successful pick. The list length, on the other hand, affects the total amount of information stored S, but not the individual amount S_i, and, hence, list length affects the probability of a successful pick, but not the probability of recovery following a successful pick. This model has only three parameters: r, the buffer size; θ, the transfer rate of information per unit of time; and n, the number of picks into memory. The data to be fit come from four different experimenters (Murdock, Postman, Deese, and Shiffrin). Presentation rates were either 1, 2, or 2.5 words per second, and list lengths were 6, 10, 11, 17, 20, 25, 30, 32, and 40; 16 different serial position curves altogether. A least-square technique was used to estimate parameters, and they were as follows: $r = 4$,

$\theta{=}0.04$, and $n{=}34$. Table I presents the theoretical and observed values; overall, the predictions are about as accurate as noise in the data would allow, the primary effects and asymptotes being accurately fit for each of the 16 conditions examined. This is especially impressive because only three parameters were estimated, because a wide range of conditions with related wide ranges in performance was examined, and because the data were collected from a number of separate sources.

These results demonstrate the workability of search mechanisms as an alternative to interference formulations, but do not, by themselves, allow us to choose one over the other. Nevertheless, there are a number of ways to experimentally demonstrate that search processes do exist in many situations. For example, the search theory assumes that on a retest some information skipped the first time will be found and reported: such is the case. We have even found, in free verbal recall, that an item not reported from a list is sometimes reported in error, as an intrusion, in recall of the following list. In any case it seems clear that a considerable array of data may require a search mechanism for explanation. Because this mechanism can also predict effects predictable by interference, the interference models tend to become redundant. I will conclude, therefore, with a prediction that the next few years will see many of the results of long-term forgetting cast in terms of appropriate search models.

TABLE I.—*Observed and Predicted Serial Position Curves for Various Free-Verbal-Recall Experiments*

List	Point 1		Point 2		Point 3		Asymptote		
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Number of points
M–20–1 ...	0.45	0.45	0.27	0.37	0.20	0.29	0.16	0.22	2
M–30–1 ...	.38	.35	.30	.28	.21	.22	.19	.17	12
M–20–2 ...	.55	.61	.42	.51	.37	.41	.31	.32	2
M–40–1 ...	.30	.29	.20	.23	.13	.18	.12	.14	22
M–25–1 ...	.38	.39	.23	.32	.21	.25	.15	.19	7
M–20–2.5 .	.72	.66	.61	.56	.45	.46	.37	.35	2
D–32–1 ...	.46	.33	.34	.27	.27	.21	.16	.16	14
P–10–1 ...	.66	.62	.42	.52	.35	.42	.34	.32	7
P–20–1 ...	.47	.45	.27	.37	.23	.29	.22	.22	17
P–30–1 ...	.41	.35	.34	.28	.27	.22	.20	.17	27
S–6–1	.71	.74	.50	.64	.57	.52	.42	.40	3
S–6–2	.82	.88	.82	.79	.65	.66	.66	.52	3
S–11–1 ...	.48	.60	.43	.50	.27	.40	.31	.31	8
S–11–2 ...	.72	.76	.55	.66	.52	.54	.47	.42	8
S–17–1 ...	.55	.49	.33	.40	.26	.32	.22	.24	14
S–17–2 ...	.68	.66	.65	.56	.67	.45	.43	.35	14

REFERENCES

1. BOWER, G. H.: Applications of a Model to Paired-Associate Learning. Psychometrika, vol. 26, 1961, pp. 255–280.
2. SPERLING, G.: The Information Available in Brief Visual Presentations. Psychol. Monographs 74 (whole no. 498), 1960.
3. AVERBACH, F.; AND CORIELL, A. S.: Short-term Memory in Vision. Bell Tech. J., 1961, pp. 309–328.
4. ESTES, W. K.; AND TAYLOR, H. A.: Visual Detection in Relation to Display Size and Redundancy of Critical Elements. Perception and Psychophysics, vol. 1, 1966, pp. 9–16.
5. NORMAN, D. A.; AND WICKELGREN, W. A.: Strength Models and Serial Position in Short-Term Recognition Memory. J. Math. Psychol., vol. 3, 1966, pp. 316–347.
6. PETERSON, L. R.; AND PETERSON, M. J.: Short-

Term Retention of Individual Verbal Items. J. Exptl. Psychol., vol. 58, 1959, pp. 193–198.

7. MILNER, BRENDA: The Memory Defect in Bilateral Hippocampal Lesions. Psychiat. Res. Reps., vol. 11, 1959, pp. 43–52.

8. HEBB, D. O.: Distinctive Features of Learning in the Higher Animal. *In:* Brain Mechanisms and Learning (J. F. Delafresnaye, ed.), Oxford University Press (London), 1961, pp. 37–46.

9. MELTON, A. W.: Implications of Short-Term Memory for a General Theory of Memory. J. Verbal Learning & Verbal Behavior, vol. 2, 1963, pp. 1–21.

10. MURDOCK, B. B., JR.: The Serial Position Effect of Free Recall. J. Exptl. Psychol., vol. 64, 1962, pp. 482–488.

11. POSTMAN, L.; AND PHILLIPS, L. W.: Short-Term Temporal Changes in Free Recall. Quart. J. Exptl. Psychol., vol. 17, 1965, pp. 132–138.

BIBLIOGRAPHY

BERNBACH, H. A.: Decision Processes in Memory. Psychol. Rev., vol. 74, 1967, pp. 462–480.

COHEN, B. H.: Recall of Categorized Word Lists. J. Exptl. Psychol., vol. 66, 1963, pp. 227–234.

HARLEY, W. F., JR.: The Effect of Monetary Incentive in Paired-Associate Learning Using an Absolute Method. Psychonomic Sci., vol. 3, 1965, pp. 141–142.

HUBEL, D. H.; AND WIESEL, T. N.: Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. J. Physiol., vol. 160, 1962, pp. 106–154.

KINTSCH, W.: Memory and Decision Aspects of Recognition Learning. Psychol. Rev., vol. 74, 1967, pp. 496–504.

MANDLER, G.: Organization and Memory. *In:* The Psychology of Learning and Motivation (K. W. Spence and J. T. Spence, eds.), Advan. Res. & Theory, vol. 1, 1967, pp. 328–372.

PENFIELD, W.; AND ROBERTS, L.: Speech and Brain Mechanisms. Princeton University Press (Princeton, N.J.), 1959.

POSNER, M. I.: Components of Skilled Performance. Sci., vol. 152, 1966, pp. 1712–1718.

SALTZMAN, I. J.; AND ATKINSON, R. L.: Comparison of Incidental and Intentional Learning after Different Numbers of Stimulus Presentations. Am. J. Psychol., vol. 67, 1954, pp. 521–524.

SPERRY, R. W.: Cerebral Organization and Behavior. Sci., vol. 133, 1960, pp. 1749–1757.

TULVING, E.: Subjective Organization in Free Recall of "Unrelated" Words. Psychol. Rev., vol. 69, 1962, pp. 344–354.

Comments

GEORGE N. CHATHAM

NASA Office of Advanced Research and Technology

GEORGE N. CHATHAM: Most of the conference thus far has involved problems in individual decisions, decisions made by an individual in a tight situation. Long-term space flight is not in that category, yet it is what you might call at this stage a hypothetical mission. As a hypothetical mission, it is an important analytical tool for decisionmaking of a different type. The whole category of decisions for long-term space flight could be considered an applied category, because there is an object in mind. But, of course, such flights have a whole hierarchy of problems associated with them, some of which are applied and some of which are not at this stage.

The kinds of decisions that enter into long-term space flight are such decisions as when should we actually start such a venture? Which flight should precede or which should be the next? One must consider the feasibility of the flight in terms of its mechanics, what contribution it will make to science, to space-flight problems, to national requirements, or national problems. The chances are that the space program is going to be twice as old as it is now before the decision is actually made to begin work for a long-term space flight. But what class of problems must be solved for such a flight? The problems must be anticipated as best we can, because obviously they are not going to be answered in the final form.

Why would NASA be interested in research in this area? Because they want to be prepared and because they want to improve the second flight, which means they must know what went on in the first one.

What is needed for correct decisions regarding the formation of these long-term missions? First, we must provide the facilities that will make correct decisions possible. Second, we must make sure that it is possible for the astronaut to make the correct decisions once he has the facilities. The first category involves problems in architecture, which is a relatively new thought in NASA. There is only one very small effort involved on what kind of architecture you can do in space. It is a completely different concept in architecture, in which the central purpose is to maximize the use of volume and surfaces that are available within the weight limitations. They are studying ways to provide maximum utility for the finite amount of space for 7 to 12 men. Whatever number is eventually decided upon, the architectural principles will remain fairly constant.

The first category also raises questions with regard to the equipment. I think that any equipment should be as fail-safe as possible. But it should also provide time for decisions; that is, it must maximize the time available for decisions.

There is also the matter of the type and amount of communication that you wish to provide, within the ship and also to Earth. To have complete ground control for such a long-range flight would be fairly difficult.

Medical provision for emergencies on board, both of a physical and mental type, is another very important area that is some-

times omitted. Also, what do you do in the event of death? These decisions must be made before beginning final design of the ship. All of these items are in what you might call the facilities area, the mechanics of the spaceship itself.

In the human area, one of the main problems is that of selecting persons who would be suitable for such a trip. This is an area that has been investigated a great deal and the findings are very contradictory. It is a quagmire of opinion at the present time. Cross-training is, of course, essential; there are only a few men on board and the dangers are not completely foreseeable. Preferably, every man should know every other man's job well enough to do it, not only in terms of running the craft, but also in terms of giving medical treatment or the carrying-out procedures that are called for in the book of ship's rules.

Another major question is what kind of social structure shall be provided? Obviously there is going to be a book of ship's rules, some of which are obvious and can be foreseen readily, others of which we will have to make best estimates on, such as the degree of authority that is present on the ship. Also ship's rules should spell out the use of time, on duty as well as off.

These are the main areas of concern. They contain problems, many of which we have talked about. Decisions will have to be made in these problem areas prior to the initiation of long-term space flight. The important message is this: there is a hier- archy of these decisions. People who are wondering what areas would be most relevant to work in have to think in terms of this hierarchy. They must consider whether they are really worrying about the fifth decimal place or whether they are making a contribution that will be actually needed.

In reviewing the research in this area, I have found that a great deal of what is done under the label of long-term space flight is simply teasing a decimal place that no one is particularly interested in. This is unfortunate when there are many glaring problems facing us.

The persons interested in getting into this area and working with the Government on such projects would have to consider first of all what kind of hierarchical structure the problems have. The writer of a proposal has to think in these terms: what problems exist, and what is their relative importance? And, of course, what is the commonality, which is one of the things I think is most important in this meeting. The answers to many of these problems relate to many problems. An answer might have been developed for a specific problem, but, of course, generality is an important aspect of it.

This paper was initially supposed to be a review of the research on long-term space flight presented here. There was not really enough aimed at that problem to review. What I have done is to capsule the whole problem in terms of how a planner would view it at this time.

Panel Discussion

STANLEY DEUTSCH, *Chairman*

NASA Office of Advanced Research and Technology

EDWARD M. HUFF: I would like to ask your views about the capabilities of the astronauts on long-term flights. There seems to be some contradiction in strategy between having scientist-astronauts and having a meaningful redundancy capability.

GEORGE N. CHATHAM: The ability of each astronaut to do the other's job refers to their versatility in the maintenance and running of the ship as well as their ability to provide the medical treatment required to maintain each other. In order to establish the scientific training they would need, we must first define the purpose of a long mission. For example, are we to bring back 50 pounds of Mars, as we are doing from the Moon? Once this task is determined, perhaps the category of the personnel on board can be more clearly examined. Whether or not a scientist is needed, as opposed to a test pilot, is a question that will have to be settled then. It certainly is not settled now. I have no hint of it. Do you?

R. MARK PATTON: No; except that my impression is that the space program will tend to go the way it is going unless something drastic happens. I have read from time to time that someone is proposing that the typical present astronaut will not be able to make it because of personality or temperament. The proposals can be as extreme as sending a legless Buddhist monk to save weight, and at the same time get someone who can stand the psychological stresses. I presume that people like the present pilot astronauts will, in fact, go on

any Mars trip. Any realistic planning must presume this.

CHATHAM: You can take the other side, too, and say: Is the scientist such a beautifully balanced personality that he could make it when the typical astronaut could not?

PATTON: The one thing we have been thinking about, but are not really far enough along with, is that we may find differences according to the motivation. It seems to me that the trip to and from Mars might be very exciting for the scientist-astronaut, particularly if he is carrying out his experimental activities all along the way. But the trip might turn out after a while to be very tedious and dull for the pilot-astronaut. It might be very tedious for him to monitor never-occurring warning lights for system failures. So I see a possible dichotomy. We have been interested in this from a laboratory approach, because we thought we could manipulate the effects of inherent job interest on various things that occur in groups, the interpersonal relationships. Virtually no research has been done in this area. We have some ideas along the line, but they are not easy to implement.

Research in the area of team performance more commonly concerns itself with such things as leadership roles and coalition formation. The question of who forms a coalition in a mixed pilot-scientist group should be a matter of interest.

CHATHAM: There has been work in this area. The armed services have had a problem in small group behavior for a long

time and have written a great deal about it, some of which is useful. This is currently being surveyed by Dr. Sells at Texas Christian University, under a study Deutsch is managing. Part of it is already out, is it not?

STANLEY DEUTSCH: Yes. To go back to Huff's question, basically the redundancy will be in terms of the operational, the maintenance, the command structure, or the station-keeping activities. I do not feel that we will be able to afford redundancy in terms of scientific capabilities. If we get the scientist on board, the men will represent separate technical specialties. And if, indeed, you wipe out part of your scientific program, this is not as catastrophic as wiping out part of your operational program, which could mean the safety of the crew.

WILLIAM ALLEN:[1] The results of Mission Analysis Division maintainability studies, matching the crew with the reliabilities of the system, show that you must have maintenance people available at all times; therefore, a scientist—until you get a crew of 12 or more—will have to have some maintenance skill.

CHATHAM: There will be no room, I assume, at this time for a prima donna on board.

ALLEN: He can be a prima donna, but he must also be a good electrician.

PATTON: My theory has always been, not a good electrician, but an electronic genius, if I were going along.

JOSEPH MARKOWITZ: Is there any information on pilots, with a relatively low workload on a long flight, going off autopilot and introducing motivation?

GEORGE A. RATHERT: It happens all the time. They will interrogate the system or——

STEVEN E. BELSLEY: Have you never taken an 880 from San Francisco to St. Louis?

RATHERT: It is more than possible—it is highly probable.

DEUTSCH: In one of the studies, we are trying to determine what you do with off-duty time. There has been a lot of consider-

ation to structuring the tasks for the astronauts and astronaut-scientists. Everybody says that they are going to be very busy. But for 500 to 600 days, will they continue the level of intensity of effort? The question then is what kind of off-duty or discretionary activities should be provided (I say discretionary, not programed, not forced) that will permit them to avoid the problems mentioned?

MARKOWITZ: I have one other question. Over such a long time period, say, a 2-year mission, may people lose interest in their own tasks and get interested in tasks that were not originally assigned to them? The scientist being most interested in flying the craft manually and the pilot being quite interested, for example, in teaching him? And at the end of the mission they might have quite different roles; that is, the pilot may become very interested in the scientific observations.

DEUTSCH: Actually this strikes me as something that should be encouraged, provided that the scientist remains the scientist and the engineer does not become the scientist solely because he has a strong personality. These are the kinds of problems, I think, Patton is referring to.

MARKOWITZ: As long as people's skills are rank ordered according to their command responsibility, that would be all right. Suppose, however, that the scientist ends up being a better pilot than the pilot, or vice versa?

PATTON: I have heard people state that it is easier to train a pilot to be a scientist than it is to train a scientist to be a pilot.

WARD EDWARDS: We have several scientist-pilots in the audience. They might disagree.

MARKOWITZ: My last question is whether the pilot's performance may deteriorate as he increases his specialty in some other field?

RATHERT: I would say the probability of that occurring is extremely small.

DEUTSCH: Constant refresher training, perhaps.

MARKOWITZ: You are not going to fly any onboard simulations?

---

[1] NASA OART Mission Analysis Division.

HAROLD G. MILLER: They had better plan to.

DEUTSCH: How would you do that? With real-type problems? If you give meaningless tasks, you will have problems. In agreement with Miller, we ought to try to provide refresher training. The nature of this refresher training is, however, still a good question.

MILLER: This is going to impact spacecraft design and computer size.

RATHERT: You are discussing a problem that has been with us for a long, long time. I have never met a test pilot who did not think he was a better airplane designer than I am; and I am sure that most of the test pilots think I act like I know more about flying than they do. There is conflict in your own field of psychology about the design of instrument displays. This has been going along forever—the pilots claiming you are trying to tell them how to fly the airplane, and you thinking that they are encroaching on your field and telling you how to solve information-processing problems.

CHATHAM: I agree with you fully. You do not have to be ordained to be a scientist any more than you do to be a pilot. It is just a matter of the application of certain information or skills in any case or any profession.

DOUWE B. YNTEMA: You draw a rather dark picture about research in decisionmaking in this area. At least if we accept what Edwards was urging: that discussion of decisionmaking must involve the question, "For what purpose are these decisions being made?"

CHATHAM: The purposes are also selected by decisions. You must consider what it is that would be the biggest payoff on the first flight. This, too, is a decision.

RATHERT: I think you are very, very conservative in your estimates of the motivation on the very long flights.

We put a test pilot and a scientist together in an Apollo mockup, 65 cubic feet per man. We did a 7-day experiment. We were told before we started that we would never get a serious NASA research test pilot to sit still for this kind of confinement. He walked out of our test, flew across the country, and went through the same experience again for Martin Marietta voluntarily. The motivation is extremely high and extremely untapped. I hear an awful lot of what I must frankly call hot air, people expressing doubts about the motivation. The motivation for actual mission is there, and we are going to have to stop worrying about it.

YNTEMA: There is a good deal of practical experience with a slightly larger volume of space per man and a slightly shorter period of time (like 8 months in the South Polar station, with a community of scientists), except that some of the variables are a couple of decibels off.

RATHERT: I always remember our experience with the astronauts and our own test pilots when we were determining minimum habitability space for the lunar mission. A lot of elaborate tests were made and a lot of money was spent, and the basic pilot answer was the one you gave at the start: "If you are selecting me as the astronaut for the first lunar mission, I will go in a telephone booth." And they showed they could do it.

MELVIN SADOFF:[2] I would like to react to Huff's original question with respect to whether you would have a scientist aboard or a test pilot aboard. It seems to me that we would have an awfully difficult time selling any trip to Mars, Venus, or any interplanetary flight just because we have capability to get there (like climbing Mount Everest because it is there). A multibillion-dollar program will require scientists on board. Many documents available in NASA publications say just exactly this; namely, the reason we are going is for scientific inquiry.

Now, the extent to which test pilots and scientists will be combined is, perhaps, the question that needs to be included in the decisionmaking processes involved in space mission planning.

CHATHAM: Actually when words like "scientist" are used, ordinarily one thinks that reference is made to a specific person.

[2] Chief, Man-Machine Integration Branch, NASA Ames Research Center.

More properly, a scientist is a person conducting scientific work. No matter whom you train to do this, at the time he is working as a scientist, he is a scientist. These points must be considered, too, when you are thinking about who will go.

BELSLEY: This argument that you are discussing right now has been solved, at least at one level, by saying that the scientists do not believe that astronauts can become scientists, so they insist on scientists becoming astronauts. This is the position the Space Science Board, or whatever, took to force the inclusion of scientists in the cadre for going to the Moon.

JOHN W. SENDERS: In the 1962 summer study at Iowa, this problem, as I recall, was aired for the first time. The document that came out of the study specified a hierarchy of differential degrees of scientism and astronautism, culminating with the conceivable notion that there might be situations in which people who had no astronaut capability whatsoever might be sent along as pure scientist passengers. This was projected for sometime in the 1970's, but nonetheless it still remains a possibility.

LLOYD A. JEFFRESS: The astronaut must make decisions every once in a while about which of two lines of action to take, and it seems to me the same thing is going to be true of the scientist. Unless a person has a fairly rich background on which to base the decisionmaking, he is not going to be as useful a member of the crew as if he were merely trained to do certain things. Maybe these things are not practical and something totally different will turn out to be the thing that should have been done.

HUFF: Along these lines, exactly what kind of conflict situations could exist between the scientific activities and maintenance activities? If one allows for the fact that various problems can arise at any time, then, depending on the scientific workload and the necessity of using scientists for maintenance purposes, you get into a decisional bind as to what should be sacrificed for what. Not all emergency situations have catastrophic impact. Some could be postponed and delayed.

Really, as I see it, it comes down to a matter of workload.

CHATHAM: It does come down to a matter of workload, but it comes to something else too, in terms of which action comes first when neither would be catastrophic. This goes back to another problem: the degree of authority on board, the kind of social structure established on board. If the captain decides which action comes first, even when no emergency exists, the social structure is highly authoritarian. On the other hand, are we going to leave it up to a vote? This will have to be settled.

PATTON: Chatham, do you not think this problem is almost directly analogous to the Antarctic situation in which, as I understand it, there is a military-scientist interface? The military has the authority over the base and the military functions, and the scientists have the authority over their science. It seems to me that there is a direct analogy here that would suggest you do the same thing in space flight. I cannot imagine setting up a really democratic space flight society.

CHATHAM: I do not think anyone can. There is no such precedent in any ship that ever existed.

PATTON: The commander is going to be the commander. And, if in his judgment mission safety is endangered by anything, he is going to make the decision that it must be the other way.

CHATHAM: I do not think any commander would attempt to control the scientist to the degree of telling him how to conduct his experiment. His concern should be the relationship of the experiment to the overall mission in terms of safety, the use of facilities, or whatever.

DEUTSCH: If current plans go through, we may see a study that will take another tack. Chatham indicated the fact that he cannot think of any mission, certainly any real-role mission and probably simulated missions as well, where command structure was not formalized. But we are thinking of the possibility of an undersea habitability study of four men in a habitat 50 feet below

the surface of the water for some 60 days. They will be monitored both physiologically and psychologically. This may give us another slant on the question you raised: Is the only reason that a commander will be designated because he has had prior experience in a study of this nature? The subjects will not know each other and will not have lived together for very long before they go under water. In addition to this, there is no reason to believe that the predesignated commander is the best qualified man, except by virtue of experience, to head up this activity. In some of the prior Sealab studies, the most experienced diver has become the functional commander. Maybe this will carry over. We do not know. But this is a case where there is not a highly formalized, structured system. It is deliberately set up this way from the social dynamic standpoint to see what happens in the command structure.

PATTON: I am reminded of the Admiral Hornblower books, in which the admiral must defer in certain matters to the captain of whatever ship he happens to be on. Even though Admiral Hornblower really might know more about sailing the ship, it would be the most gross, inexcusable violation of etiquette of the situation for him to tell the captain anything about how to trim the sails, or whether the wind is coming up, and so forth. I think this still goes on because there are admirals aboard Navy ships today who do not tell a captain how to run his ship.

DEUTSCH: The captain is responsible for the safety of the ship. If anything happens to the ship, it is the captain, not the admiral, who will get roasted. This may be why the admiral does not interfere.

BELSLEY: They tell him where to go, but they do not tell him how to get there.

RATHERT: You are overlooking one conflict. When you begin to get the research results of Serendipity on the mission, some of the conflicts between the scientists will make the conflicts between the scientist and the pilot look like nothing.

YNTEMA: I understand that in the little 12-man station close to the South Pole, the command structure was interesting and somewhat flexible and might well be worth interviews. There has also been a good deal of experience with long research voyages. I gather that there is a rule book, but the actual decisionmaking processes are rather subtle.

CHATHAM: That is an excellent point. There is another too, that backs your statement that there will be squabbles among the scientists. A good portion of the Antarctic study showed that things become more important as the number of available stimuli decrease. In other words, things become exaggerated. This seems to be a uniform observation of people in small isolated groups: the problem of exaggeration, not only of things to do but also of irritations. We should anticipate squabbles, and should provide ways of settling them.

RATHERT: Ames has its Convair-990 flying laboratory, a large 4-engine jet airliner that carries typically 8 to 10 scientific experiments (aurora observations, comets, eclipses, this sort of thing). There are 8 to 10 investigators, typically rather senior men from the universities. Ames sends along on this airplane a senior scientist, a peer of the investigators, who is there for the obvious reason and appears to be a necessary part of the crew. These are only 8- to 10-day expeditions, but it seems necessary even there to provide a peer judgment. In other words, the pilot of the airplane conceivably could act as the captain in determining how to position the airplane, but it is actually done by the process of having a peer of the investigators on board.

PATTON: Does he have authority over the pilot?

RATHERT: He has authority to tell the pilot what to do, until, in the judgment of the pilot, he is endangering the ship.

RICHARD C. ATKINSON: I wonder about this concept of conceiving of the crew as the onboard personnel. Given modern communication systems and all, why must one conceive of the crew as people sitting on board the ship? The executive officer on the bridge of the ship makes instantaneous decisions without consulting the captain—if they are

called for. Why not redefine the concept of the crew as certain specialized people sitting on the Earth plus those in flight? Frankly, I am not much impressed with the idea of sending a scientist off in space. The only rationale, from my point of view, in sending a scientist is not instantaneous decisionmaking, but his ability to communicate the scientific data to an Earth station. I wonder if there is not something to be said for chucking this idea of onboard personnel and redefining the concept of the crew as a limited set of people sitting on Earth. That does not mean everybody at Houston Center.

MARKOWITZ: There is almost this idea now. The same argument you advanced for sending a scientist up was the reason that Cap-Com has been an astronaut.

PATTON: Long communications delays are a factor. On the Venus swingby, a 30-minute delay for getting scientific advice from Earth would be intolerable. A lot of scientific decisions will have to be made depending on the state of the equipment at the time the ship goes into the swingby. There has to be the competence on board to make a decision, whether the man is labeled a scientist or a scientist-astronaut.

MILLER: The inertia of the system is such that the flight director and mission director and people on the ground say sort of when to come home over the objections of the crew, and the crew maybe acts like a captain; but in major directional decisions the ground has the final say. I would see no reason for this to change in the longer flights.

PATTON: They are acting like the admiral, are they not?

MILLER: Yes.

DEUTSCH: Is that not like the story about the man who says that he and his wife have decided she is going to make the minor decisions and he is going to make the major decisions, and so far there have been no major decisions to be made?

MILLER: No, that is not the case. In fact, history will bear me out. On the GT-8, where the people on the ground told the crew to come down because they had violated the mission rule of using the reentry attitude

control gas and broken the seal, they said, "If you do that, come back, because it could leak out." The crew did not object, but they would have liked to stay up a little longer.

DEUTSCH: Let me take the chairman's prerogative and cut off this line of discussion.

I want to indicate a feeling of frustration on my part in this respect: There were a number of problems identified, and I feel that we are going to have another session in which those problems identified become the topic. I think it is typical of meetings of this nature that you usually raise more problems than you answer. Along these lines, then, I am not certain of the rules of the game that permitted discussion here of some of the research that I heard described. I am not sure how these discussions fall within the rules of the game of decisionmaking. I feel intuitively that the studies could move in that direction, but I do not have a satisfactory feeling of closure that the applications were made in terms of the name of the game for decisionmaking concepts, even with the three categories that Yntema described.

We know that if a man is on line, he can respond more rapidly than if he is off line; he has to get on line to find out what past history has just evolved that may have led to the emergency and what steps he can take to satisfactorily resolve the problem. Should an astronaut always be on line, to what extent, and within what time frame must he operate?

I am not sure that the discussions of the research really fit into providing answers to some of these problems. Does anybody have any suggestions as to how these types of research can provide the kinds of answers for some of the problems raised?

ALLUISI: I want to go back to what Birdsall was saying. I would say that the question you just asked and the approach you took to it is fully an approach of a researcher and not a development man, in the sense that we would always like to know more than we now know when we build a system. When we get ready to build a system, we have to build it on what we have and know. We have to make our estimates; we have to make

our guesses; we have to make them the best way we can. We very often know of areas in which we do not have good research backing up the design, but in many cases we do not even know we lack the research. But this is not peculiar to psychology. It is generally true with every field of knowledge in the world.

The problem is how to get the research kind of information into the form that the applications person, the actual designer, can use. I do not know that we will ever have a solution. It is like a different language.

I am not even sure we can solve our problem if we put people in the middle. What would the person in the middle do? We are hunting for a solution in which we ask for a man to be the Renaissance man, the jack-of-all-trades who knows all ends. Most of us do not know enough about either the research or the application ends. There is too much on both sides for any one person to know. All the middle man can do is be a catalyst. All he can do is try to bring together the people with the knowledge on the two sides.

I have done research for a number of years on reaction time that is directly related to decisionmaking, specifically showing that reaction time is a linear function of the number of alternatives in a choice reaction task, and showing that the rate of gain of information is a constant. This research also shows that the rate of gain of information may be varied by changing some aspect of the performance situation, such as the stimulus-response compatibility. I show that if stimulus-response compatibility is high, as it is in a compact console, the rate of gain of information will be a constant, but it will be at a lower slope than when the stimulus-response compatibility is low. Mainly, what this says is that with training, the stimulus-response compatibility will be increased, and decisions may enter the class of behavior that is essentially automatic. Decisions will be made more rapidly, and may even become independent of increases in the number of alternatives. What does this mean? As we learn things well enough—as we overlearn

them—we can handle a wider variety of things with essentially a flat reaction time.

Now you ask me to apply that. How do I apply that in the design of a system?

CHATHAM: You must give the engineer the criteria that he needs. The engineer does not have these criteria. You said a compatible ensemble flattens this curve. The engineer must be given the parameters that you know will make that ensemble compatible, otherwise he is going to spread it out to suit his own convenience.

ALLUISI: We are never going to find that we can have a large body of researchers who are doing exactly what the design engineers would like to have them doing, nor can we have a group of design engineers who are doing exactly what the researchers would like them to be doing.

We have different goals to achieve. And we are going to keep approaching our goals in our own best way. What we need to do is to come together like this more often and come together a little less formally than we have, not merely with some people stating problems and some people stating research, but rather with some of the people who are stating research sitting down with some of the people stating problems and the two of them jointly looking at each from the other's point of view. I would like to get some of the design people looking at the research, because they will assimilate it and they will be applying it without even knowing what the rules are.

Again, I want to stress that I do not think this is peculiar to our field. I think it is the same in every field. I think you might have a very fine physicist who is the world's expert on the tensile strength of metals as a function of temperature but who cannot design a bridge.

DEUTSCH: I did not want to give the impression that I felt that the research described was useless and meaningless. What I did mean to imply was that certain problems were identified. You cannot provide all the problems in advance to the speakers and say, "All right, here are the problems."

ALLUISI: I do not shy away from the

application. I think any scientist who shies away from the application is that much less a scientist and more of a theologian, because the test of any science is the technology. If it does not work, it is not right.

DEUTSCH: I would agree. I think you gave some very pertinent examples. Again I did not mean to imply that the application was not there. But how does control theory tie into the decisionmaking problems that were expressed here?

JOHN A. SWETS: May I speak to that point? You are asking about the organization of the conference, and I protested on the first day that we did not start with a topic and then choose speakers. If we had done that and started with decisionmaking, we might have had any number of people. In fact, we started with speakers, with people doing research under Ames support, and then probably did a poor job of choosing a title rubric to bring these people together. I protested that maybe human information processing might have been a more appropriate term that would include memory, attention, and a few other things; manual control is another. But even if we worry about decisionmaking, it depends in many instances on memory, on attention; in many cases it is implemented through a manual control process.

EDWARDS: I thought, initially, when I heard what was going to go on here, that the meeting was misnamed, but found myself being argued out of it by the speakers I listened to, even though they were talking about things like manual control. Perhaps the key to the arguing-out process was the point made first by Yntema and then made again by a number of us that a convergence is visible in trends in this kind of research. I think that perhaps your concern illustrates this particularly well. It is increasingly clear that the researchers on manual control are really looking for the places where values of various kinds get in, where decisions really do enter the process.

Signal detectability theory is a clear-cut attempt to use decision theoretical ideas in what was traditionally an area of research not thought of in this way. A great deal of the motivation for that analysis was to give an account of decision effects so that sensory aspects of the total system can be studied without contamination by decision aspects. But inevitably the decision aspects have been studied too. About the only traditional area of human experimental psychology research that is somewhat resistant to this ingress of decisions is verbal learning, and even there the signal detectability ideas are being fruitfully applied.

So I was argued out of the feeling that a lot of this conference did not have much to do with decisionmaking, essentially because at the abstract level it seems to become increasingly clear that you cannot talk about any kind of human intellectual activity or human purposive activity without introducing decision-theoretical concepts.

DEUTSCH: I think somewhere you must define the limits. It is like saying that any research that you do in any of the areas is applicable across the board. I suppose to an extent they are.

EDWARDS: My point is that this recognition of the pervasiveness of decisions is more than lipservice. The lipservice could have been available 20 years ago because, as a sort of abstract point, it was clear 20 years ago that actions are the result of decisions. But, increasingly, this is coming to be not only obvious but theoretically useful.

JEROME I. ELKIND: I disagree with the prior statement. It seems to me that one of the beauties of this theory is that the limits are not clear. You do not want to define the limits; you want to extend the range of applications.

One main reason is the kinds of systems with which you are concerned. If you can find a common basis for talking about many of them or most of them, you have a way of starting to deal with them.

SWETS: You stated one problem—if you let the machine make the decisions and the human operator dope off, then he does not do very well when he has to override the system, the automatic decisionmaker—and two or three others of that sort: I suspect

that if we had invited people to come to speak precisely on those problems, these people would not be here. In fact, it is nice to have researchers here, because you want them to start thinking about some of those problems.

DEUTSCH: This is why I indicated that this is perhaps not the last meeting. Perhaps the problems that were thrown out will provide grist for the mill, if you will, and provide indications of things that might be looked at that are of direct interest to those who are working in these research areas.

SWETS: We did not hope the first meeting would be the last meeting. I guess we do not think we are that far along.

THEODORE G. BIRDSALL: Let me add one point along with Edwards'. I think that a lot of the human-performance measurement reported shows that the abstract models or just mathematical models or the models the engineers use for decisionmaking in equipment can indeed be applied to human beings. You can begin to quantify a lot of it and it can be talked about in the same terms.

If we talk about human decisionmaking or machine decisionmaking in the same terms, then we can build equipment to replace the human. Yet, the kind of decisionmaking that almost always comes up, and, I think, the kind that requires the human to be there is the override-type decisionmaking on the things you do not plan for, the things that you cannot take into account when you do build the equipment. In Edwards' examples, the decisions were made on the basis of facts that were there, but that were not in the original model. The original model for the decisionmaking did not take them into account. That is one problem with hardware. You design it and it has limitations. If you tell it to forget about something, it beautifully forgets it. A lot of the human's job is to keep track of all the information inputs that really must be thrown in all the time so that he can override the machine on those things for which the machine is just too stupid to have remembered had happened or was never programed to even realize that those things might be relevant.

That is a hard line to draw. If we thor-oughly understood everything that was going to go in, we could automate the whole process. We must realize that we cannot. So we put a tremendous burden on the human being. He has to be the supergenius who is better than the equipment designer who takes into account all the things that might happen, all the very small, low-probability things for which the machine designer said, "Well, we will just take those out." If you want to automate it completely, how would you ever build the automatic equipment that is going to take this into account? How would you feed this information into it? How much storage, how much capacity, how much real computation of all these possibilities would it have to account for? I think somehow working down this line will be the kind of research that will pay off in what the human part in this decision is going to be.

ATKINSON: I get the feeling from some of the remarks you have made that the feeling is that if the scientists would just really take some time and get familiar with these problems, they would be able to come up with a lot of clear answers as to how to proceed.

I am really aghast at behavioral scientists looking at a problem area and coming up with very clear principles as to how to proceed. Look at Skinner's suggestions as to how to write a programed text. It is really frightening to think of someone turning to the curriculum community and saying, "Look, you cannot possibly understand the psychological research underlying all these principles, but if you proceed in the following way you will design a good curriculum." It is just utter nonsense. What we are going to provide is worse than providing nothing at all.

DEUTSCH: I am not sure that is totally true, insofar as there are all levels of scientists and all levels of engineers. And there are those who are involved in purely basic research and those in highly applied research. In this case you do have those who fill a gap. Perhaps the area of human factors is one that is really applied engineering, applied psychology, what have you. It

does attempt to fill this gap because its workers can speak, to a certain extent, both languages. I agree that if you do not keep providing the raw material at the research end, the development work cannot continue. On the other hand, if we prepare a bunch of reports and put them in a file and forget about them, we cannot expect some practitioner to come along and pick them up and utilize them. I think there is a necessary middle of the box, if you will.

Along the lines Birdsall brought up, I would agree, how do we make these decision-making processes available, in this case, to our population, our customers, the astronauts? I am thinking in terms of the problem of training in decisionmaking. What can we do in this area, for example, to provide training for astronauts, and in what areas and to what extent would this be advisable?

EDWARDS: Those are two quite different questions, of course. You can do a lot about training once you are committed to a particular decisionmaking process. But whether it is possible to reach a firm enough conclusion about how you want the process structured to permit a training program—that is another question. And that is something about which a lot of research needs to be done.

I believe that explicit decision processes are better than intuitive ones, but that is only a bias. And I do not know of any data that satisfactorily establish that principle, even in the laboratory, much less in more nearly operational situations. I think maybe one of the things we can do in settings like MCC is to try to get some inputs relevant to questions like: Do you want to train these guys to make decisions explicitly, and if so, how?

DEUTSCH: In a sense, what you are saying is that they can be trained in decision-making intuitively. So you believe this?

EDWARDS: No question about it.

DEUTSCH: How you do it is another question, however.

EDWARDS: I am saying, sure, I know very well how to train people in formal proce-dures for decisionmaking. What I do not know is whether it is better to train them in formal procedures or just let them make decisions the way they do now.

RATHERT: At this exact point I am standing right in the middle. The X airlines' training man is saying, "How do I train pilots to handle the decision when they enter turbulent air?" How do I get you two people together?

EDWARDS: I am not hard to find.

RATHERT: In this case I have found you, but the problem in general is: What is the next move? This is an example of the practical everyday problem that walks into my office, where I am in the middle and see the development man with his problem and not the slightest idea of how to tackle it, and see the research man who thinks he might know how if he knew more about the problem. You cannot depend on serendipity to find both ends.

PATTON: It seems to me that the role of Government people in the matter is twofold. There is, of course, our in-house work, where we continually attempt to bring the real world and the laboratory together in a meaningful way. But the matter of research contracts and grants is equally important. In evaluating proposals and making decisions on funding, we are very conscious of this factor. In my experience, very few proposals show a good grasp of a real-world problem, as well as a promising scientific approach to its solution. I do not think that covering both sides is too much to ask, because occasionally it does happen.

On the positive side, I think that there is a growing awareness of the need. Really, it is the primary reason for this conference. I do not expect that we will go away with some final solution. Our idea was to provide an opportunity to do some interacting, and hopefully to bring the two domains a little closer together.

ALLUISI: It is a big part of what DOD is trying to get at in its THEMIS program, in what they refer to as the coupling mechanism. That has been our problem.

PATTON: If you are at a university, and

want your work to be meaningful and to have an impact on system development, you ought to spend your sabbatical year as a member of a systems development team in industry. Then when you return to your laboratory, you should have a wealth of information that will make your work more realistic. The alternatives are to gather the information on an occasional basis, or to hire people (team members or consultants) who know about the problems, and can educate the others. I sometimes think that people expect some formula to be developed that will allow one to do system-related research without investing a lot of time and energy in learning about systems. The fact is that you do have to learn about systems and thus become expert in two domains, which few people are willing to do.

BELSLEY: Is not the problem one of trying to couch the terms of your research, done at whatever level you want to do it, such that the man who has a practical problem can recognize that there is a body of knowledge that could be applied? If he could recognize it, he would go out of his way to apply it. The basic trouble is that there is a fuzzy area in which one guy talks in one language and the other guy is looking for something in another language. You can do this by writing summary papers in which you try to do things in a broader sense. If the information is put forth so that the development people can recognize it, they will use it. They are not stupid.

SWETS: Do not overlook Birdsall's point in this respect. Real progress has been made in describing human performance in the same terms that machine performance has been described. That ought to be couching it in language that the development people can use.

BELSLEY: I am not saying that it is not. I am saying that you must put it so that they can recognize that there is something there and they will use it. But, if the researcher insists that he is just going to communicate with himself, then we get back to the theology business.

DEUTSCH: I recall a meeting held in Washington a couple of years ago on computer augmentation of human reasoning—which sounds like it would be very much allied with this type of program. And yet as I listened to more than 2 days of constant discussion, I heard some wonderful ideas on data retrieval systems that were quite advanced, I heard some wonderful ideas on storage techniques, and so forth. But nobody looked at the customer. The wonderful computerized methodologies provided the capability for a customer to use them, but nobody looked at human reasoning and the requirements it imposed on the computer programs. They only looked at making information available for human reasoning. I feel that that did not tie the loop properly. That part of it was left open. I looked at our title, again forgive me, "Applications of Research on Human Decisionmaking," let us say even human information processing. I think that this part has got to be tied in, tying the loop together, otherwise we take our material, but we do not achieve what I consider to be a basic human feeling or desire. We like to see results come from the things we do. If we can see some of the results of our research go into an Apollo capsule or a flight to Mars, then that is a wonderful reward system. Otherwise the reward system is that which we get from our peers, but is that sufficient to justify our efforts here?

ELKIND: I guess we all realize that the reason we are having this conference is because we have not done a very good job thus far. It is not surprising that you do not see many applications represented, only a few false or good starts. We will find out in a few years.

Edwards has tried to do something. I would not be a bit surprised to see him get discouraged in about 6 months.

PATTON: I would not be surprised to see him a mission controller in 6 more months.

ELKIND: Is it a purpose of this meeting to discuss how to bring about more applications? I think that that is reasonable. It may turn out that some ideas come forth.

SWETS: You talked about a conference

on computer augmentation of human reasoning, where nothing really was directed toward the customer. I am wondering how many conferences you know of where the customer was invited at all, let alone in such quantities as have been invited to this meeting, primarily, as I previously pointed out, to keep us honest and to educate us. It is the first step in some respects, and frankly, I do not want to be too defensive about this, but it hurts a little to make a first step and have somebody say, "What have you done for us lately?" We are looking ahead. As the title suggests, we hope to identify a few applications or a few research ideas that look close to application, but if in point of fact we do not do that but let people leave in a slightly different state than they came, then I submit that that is not a bad objective. The value of a conference like this cannot be ascertained at its conclusion. If there is going to be any real value, as Birdsall pointed out, there must be a lot of hard work between people learning about development problems and people learning about research problems. It seems to me that the value, if any, of this conference will be apparent 5 years from now at the earliest.

TRIEVE A. TANNER: One thing that should be pointed out is the fact that everybody who came to this conference did not feel obligated to do so. Some of the people here had no obligation at all to NASA and just the fact that they came indicates that they are interested in helping to solve our problems. I think that is a good thing in itself.

MILLER: May I say something from a customer's point of view? With due respect to Birdsall's comments that you need a middleman, maybe you discount us from gaining some ideas from what we sit around and listen to. Elkind's mathematics befuddled me, but he did strike a nerve when he talked about his feedback loop. It brought home a point that I put to bed years ago that was using a model of the actual system to determine what is wrong with the thing. Now I am going to go back and look at it

and see where I can apply that idea. This has happened several times.

DEUTSCH: Maybe you are suggesting, in essence, that the audience is too restricted. Swets tried to keep it as a working group; and perhaps there are people who could benefit from some of the discussions and ideas that have been thrown out, but we are limited to what we have here. I assume there will be proceedings published. By the same token, it is reaching a limited clientele. Is there, in fact, a requirement for expanding a meeting of this sort?

MARKOWITZ: You cannot argue both for more informality and a wider audience on line at the same time. I think those are incompatible.

PATTON: I would love to have had 20 people, such as Miller, sitting here just for the reason that he has spoken of as this being a success. But this is not easy. Again, it is something we will work toward if we have such a conference in the future.

RATHERT: You are talking about communication problems between development and research. This is something that the old NACA wrestled with for years and years. The problem was solved by arranging periodic conferences where the research workers got up and gave 20-minute papers and we then attempted to put the research in the context of something that could be understood from that start. The whole subject of communication between the development people and the research people is one we could talk about for hours. I am in the middle. I have that job of communication.

There are a lot of factors inhibiting the flow of information. One of them might not occur to you. In the old NACA days we did not buy anything from anyone. We were not anyone's customer, and any time anybody in the aircraft industry had a problem, they came to us. The taxpayers all contributed to make us possible and to sit there. Our job was to solve their problems.

Now we are a customer. The last thing in the world the man from North American Aviation would think of doing would be to come to me and say, "Hey, there is a problem

in aircraft control design I cannot solve, but I want some help," because that may disqualify him on the next research contract. This is just one of the many factors like this I could mention. How do you solve this communication problem?

This point I mentioned is one of the many inhibiting factors. You said you would like to have 20 Millers here. Well, this is fine, but you are going to have trouble getting them.

BELSLEY: As long as we pay their way, they will come.

RATHERT: But they will not discuss their problems frankly.

BELSLEY: It depends on whom you ask, George. They must feel that we might be able to help them. But if they feel that our only purpose in life is to tell them they do not know what they are doing——

RATHERT: We have got to get them to say that they need help.

MARKOWITZ: We have come here and made, perhaps, sacrifices in the sense of admitting what can and what cannot be applied and of recognizing our own shortcomings. And I think maybe we feel, perhaps unjustly from your point of view, that this is no more and no less than we can ask from a contractor.

ATKINSON: I do not know how Miller gets in the center of all this, but I really found his comments very interesting. And I can see myself going down there and really becoming very interested in what is going on, and maybe having some input in terms of the training problems. I think nothing could be more disastrous than for someone to think that they are going to become more applied and consequently just take the standard laboratory experiment and dress it up in some way, some artificial sense that simulates a NASA-type problem. I think that is the disturbing possibility. On the other hand, the possibility for communication is a very important one. I sometimes think the problem is that people often turn to us and say, "What is the answer?" When they really should be saying to us, "*Is* there an answer?" And if there is no answer, they should be happy with the fact that they

have now supposedly gotten the necessary scientific evaluation and some judgments and also the information that there is a gap.

RATHERT: Somebody once said that the problem is that we are asking you to climb down off the basic research perch and go down to the development area. What we are saying is that somehow the information must be communicated that you do have something that would be of help. Secondly, what you have that is of help must be expressed in an understandable and usable form.

ATKINSON: But I am afraid that that must be face-to-face contact. I do not think people can write about it. I cannot write about what may be useful to NASA.

EDWARDS: It is two different kinds of communication. It can be done, it just takes special effort.

MARKOWITZ: I talked a little bit about road signs and made what seems to many of us very trivial application of concepts of signal detectability. I could have written about the principles until I was blue in the face at as low a layman's level as you would care to have me go, as anybody would care to have me go, and few in the Bureau of Public Roads would recognize those concepts. But when I use the words "stop sign," that is different. Many are product oriented in a different way, and they do not know to look for principles. No matter how well I explain the principles or the uses of the theory, until I get to their product, they cannot appreciate it. I think it is not too pessimistic of you to say there are too many products.

EDWARDS: As I see it, you have described some of the principles that underlie that kind of communication.

MARKOWITZ: I am saying at the same time that there are too many products. I simply cannot go around showing the same sort of payoff matrix over and over again first with stop signs, then with apples, then with airplanes, then with what-have-you.

RATHERT: Maybe instead of this computer going between Russian and English it should go between technical English and English.

Conclusion

JOHN A. SWETS

Bolt, Beranek & Newman, Inc.

JOHN A. SWETS: Let me begin this concluding session by attempting a summation. I would remind you of some themes of our discussion: some problems and ideas that we have turned to on more than one occasion in the past few days.

Several people presented operational problems in need of research help. Rathert discussed a variety of problems having to do with the flight path—the attitude about it, the short- and long-term control of it. He isolated problems having to do with terrain following, weapon system operation, collision avoidance, choice to go automatic or manual, allocation of function, and so forth. Miller alluded to many problems in discussing the function of flight controllers. Chatham and Patton raised a variety of problems connected with long-term missions. Belsley pointed out that it is essential to "substitute measurements for pilot ratings." No one, he said, is close to doing this. Belsley also emphasized the problem of "quantifying reserve work capacity." People are working on this—Alluisi, for one. However, no one would assert that we are very close to solving this problem. The specific problem Belsley hammered on us was "the problem of landing under poor visibility." The pilot breaks out, and has to decide in a brief time whether to land or not. Here we seem to be coming a little bit closer to the point of applying results of behavioral research. Tanner talked about research directly related to this problem.

I would like to come back to the problem later, but let me mention now some other themes that came up.

We talked on various occasions about "task analysis" and the requirements for predicting from a battery of tests to an actual situation. Senders pointed out that, if we could really do that, then we could bypass simulation. Someone mentioned the task analysis, performed by Serendipity Associates, that disclosed that the spaceship needed a commander who was not the pilot. Miller pointed out an interesting and, perhaps, not too strange result, that is, that every time a task analysis is made for his group, it turns out that the people who are doing the job now could not possibly be doing that job. The analysis suggests that it takes at least twice the number to do what they are already doing.

Another item we discussed was the possibility of a "taxonomy of decisions." This subject received impetus when Yntema responded to Rathert's challenge. I think we remember Yntema's three categories well enough, so I will not repeat them.

Another related theme concerned the "allocation of functions between man and machine." Yntema made the point that computers are likely to be making decisions where really we would prefer what he called "the leaven of human judgment." In some of the examples we talked about, machines do about as well as humans. Yntema gave one example in connection with judging the seriousness of air-traffic conditions.

Jeffress gave another example where machines were detecting at least as well as humans—although the real import of that very nice piece of research is that if we can predict human performance on a trial-by-trial basis in a psychophysical setting, then we are getting close to understanding what the human is doing.

The question of allocation of function came up again when Edwards made a prediction that, with respect to long-term flights, future flight controllers would be people with high school, rather than college, diplomas. There would be a need for computerizing decisions because the people who could stand the tedium of controlling long-term flights would be people incapable of making some of the more complex decisions. Rathert made the interesting observation that, while the capabilities of flight controllers may be going down, the movement is in the other direction with respect to air-traffic control and the SST. The two paths might cross or they might meet, and there should be a common focus of interest in what it is the controller should do and what decisions might be made better by computers.

That led us to the subject of "training." I recall from that discussion Rathert's question of whether or not there is a discipline that will undertake to "move decisions from category 3 to category 1" in Yntema's system. The more I think about it, there may be people who can undertake to do that. Atkinson might be one of those people. His background in learning and memory, and in computer-assisted instruction, gives him the right qualifications.

We talked a little about "selection and motivation." My contribution at this point is that John Pont, who is the football coach at Indiana, chose his team by giving all the candidates a personality test on the first day of practice. He chose the members of the team according to whether or not they scored high on what he called "personality factor 16," which isolates winners. He found his winners. He put the seniors on defense, the sophomores on offense, and got to the Rose Bowl, quite to everyone's surprise.

DISCUSSION

STEVEN E. BELSLEY: But look what happened.[1]

SWETS: Another surprise.

One theme of our discussion was "inadequate data." We talked about sparse data in one-shot missions versus a lot of data in idealized settings. We talked about the difficulty of getting any data at all in a certain box or cell of Tanner's payoff matrix, down in "coffin corner" where the pilot lands when, in fact, he should not have. Clearly, there is work to be done on the very high criterion, on the strict criterion, where false alarms are not tolerated, where it is very difficult to get any data to measure the false-alarm rate. The thrust of psychophysics in the last few years has been to get the false alarms up there where you can measure them. We run into problems for which that is not a simple thing to do. The Navy in some instances claims to want false-alarm rates of 10^{-6}. Those rates are difficult to measure; clearly, we need good theory to get from where we are to where we would like to be.

Still another theme that I detected had to do with "mathematical models: their usefulness." Senders spoke of the bed of Procrustes and told us how mod-els frequently determine experiments. I thought Elkind put that in a more favorable light when he remarked that the light is rather bright under certain model lamps, and we might do well to look where the light is bright. I think modern control theory was Elkind's example.

JOHN W. SENDERS: Let us not carry that too far, though.

SWETS: I think Edwards was the man who said, "Let us not carry that too far," and commented that decision models, in particular, are not a panacea. They do not really resolve disagreement, but they do focus it. The models provide a structure in which it becomes clear on which values or probabilities you are disagreeing. My aspirations may be too low, but I think that is something of an achievement. Thinking of Edwards and thinking of mathematical models and what they can do (particularly decision models), we spent a fair amount of time talking about applying models to complex real situations. Models seem to work fairly well under simple, idealized laboratory conditions. They are very difficult to apply, as Ward was able to point out, in complex situations. There is a problem here with respect to a conference like this: what you must do first with a model is to apply it in the laboratory, to refine it enough so that

---

[1] *Ed. note:* 1968 Rose Bowl, USC 14, Indiana 3.

you can apply it in a real situation. That, of course, takes a lot of time. So if one desires that the talk be about models that have been applied in real situations, the talk must necessarily be about very old models. That is not much fun and, in fact, most of us are reluctant to talk about very old models.

We next discussed a related topic. I think Birdsall got us going on this: Does one apply research ideas by setting up communication either on a man-to-man basis or on a 20-people-around-the-conference-table basis, or does one apply research ideas by a good deal of diligent work, like 100 men intervening between one research idea and one real solution? He was a little concerned that we have spent some time talking about "research" on the one hand and "real problems" on the other hand, but he did not hear much of "real solutions."

I think we did, however, communicate to each other something about real problems. I am going to take Rathert's advice and go home and read Davies' book on handling the big jets. Someone may accept Miller's offer to take a special course in flight controlling, which he would be happy to make available to anybody here. Rathert has suggested that his simulators are an excellent base for an experiment and might also be available to people here. Certainly Miller and Edwards gave many of us a much better feel than we had before for the flight controllers' task. I remember very well that dramatic tape replaying their conversations. Birdsall's stricture suggests, by the way, that the value of a conference like this will not be apparent at its conclusion; that, if there is any considerable value, by necessity it will become apparent only at some later point.

Yntema, in reinforcing Birdsall's point, noted that we must "organize to bring ideas from the research stage into the advanced development stage." He emphasized that it takes a big organization to do that. One man to one idea to one problem is not likely to work. It may be the wrong man for the problem or for the research idea. We must find a mechanism or an organization that will support many people working between research and advanced development.

Some of us, following Belsley's lead over dinner last night, felt that the low-visibility-landing problem might be a reasonable focus for organizing to bring research ideas into contact with real problems. Is there some sense in choosing such a problem and trying to organize about it in order to get some research ideas into practice? Does anyone want to speak on that point, either to the advisability of focusing on such a project or, hopefully, with some suggestions about a mechanism?

DOUWE B. YNTEMA: Maybe we ought to ask whether that is the big problem?

SWETS: Douwe is looking at you, Belsley. He wants to check that out. Is the problem of landing— the problems surrounding landing, breakout, mini-

mum rules, and so forth—the big problem? Is that the one to which some concerted effort might well be directed?

BELSLEY: It is a problem that everybody is worrying about, at least around here, in one way or another. It is a problem I feel has not had research results, of the kind we have discussed, applied to it. Decisionmaking theory can be used in many ways and, as far as NASA is concerned, I say there are two places that it can be done. One is Mission Control Center, and—outside of trying to structure air traffic control problems, which we have not attempted, nor I think will attempt in the near future—the other problem is to determine what is going on between the time the pilot gets on his landing approach course and the time he touches down. The reason I think this is an important problem is that at the last Human Factors Symposium, held in Palo Alto last May, we had invited as one of our speakers Captain Beck, who is on the Air Line Pilots Association (ALPA) Flight Safety Committee. He, incidentally, flies the Atlantic route, but he had come to Ames before and talked to us about the breakout problem. He came to give us a talk scheduled for 45 minutes. We had to drag him, screaming, from the podium after 1¼ hours. But his message was very straightforward. It was that when a pilot breaks out of the present minimums, he does not know whether he has time to make a decision on whether he should land or not. If he finds out that his instruments have been slightly off and he had to make the correction, then he must be able to ascertain that fact immediately and to do something about it. In the change from going on instruments to going on visual and coming back on instruments again, sufficient time has elapsed to compromise the entire landing situation. His plea was: Is there a straightforward way to assess the problem so that a pilot in his position, representing ALPA and also flying, can make a rational determination of what some of these rules ought to be? This is a perfect example in which you should apply decision theory to establishing these rules. It is like the mission rules. They are establishing them and carrying them out, all in one organization. The aircraft-landing situation has placed the manufacturers, ALPA (which is the union and has certain methods by which it can convince the others), and the airlines in a triumvirate that argues among itself. You need some data. It seems to me that the most straightforward way of getting some reasonable data is to apply this kind of research to that problem.

I am not saying it is the most important problem in the world. I say it is an important problem and nobody is doing anything on it. It is going to get more important as time goes on.

YNTEMA: That answers that question. The second question that someone who had struggled with human factors work over many years asked is: How

many years do you have to solve this problem before it becomes obsolete?

BELSLEY: I do not know how to answer that question, because I do not think it is going to become obsolete.

I do not think that the situation is going to be such that there will be an automatic system in this country for a long period of time. If you turn over the landing operation to an automatic system, then that puts a cutoff on the thing. I cannot see ALPA doing this.

YNTEMA: Suppose it takes 3 years from now to get results into the field. Is that still useful?

GEORGE A. RATHERT: Even if you go to an automatic system, you will still have the human monitoring problem. The man needs a set of criteria and a set of mission rules.

BELSLEY: No; you will not. It is either all one way or not. And you are not going to be able to monitor, get that out of your mind. If you go automatic, the pilot does not do anything. There is a point at which he cannot do anything. You cannot expect him to do anything.

RATHERT: Let the record show I disagree.

BELSLEY: That is why they have not solved it. The British are making landings on automatic and, if they try to pull off and go around when they are almost at touchdown, they have had it.

JOSEPH MARKOWITZ: That is caused by the dynamics of the craft, not the speed of the decisionmaking.

LLOYD A. JEFFRESS: I am thinking vaguely of something called GCA.

BELSLEY: The GCA works, but does what the Mission Control Center does. It transfers control and command of the aircraft out of the cockpit onto the ground. So you must convince the ALPA that that is where they want their control, and then they will transfer it. But they have not done it yet, have they? They have had GCA since the end of World War II, and when certain people use it, it works like a charm. You know that; everybody knows it.

MELVIN SADOFF: Addressing myself to the first question—whether the specific category II problem is a real one and will be with us for several years— I do not think there is any question about that. I think, further, that you can consider the category II problem as one of a spectrum of problems related to landing in reduced visibility, from category II down to category III. There will be a spectrum of configurations starting with the current way of handling approaches—which is partly manual and partly automatic—to a completely automatic system. Perhaps for the SST, 8 or 10 years from now, we will have completely automated and reliable zero-zero landing capability. I want to point out that the decision problem is a significant part of the overall problem, but it is not the only one. In other words, if you are successful in applying the talent available

in the decision-theory field to this particular problem, that will not be the complete breakthrough in solving the whole problem. It will be one important element, though. A number of others are completely outside the field of decision problems.

JEROME I. ELKIND: What are some of the other principal problems?

SADOFF: According to Captain Beck of TWA, a change in display configuration may obviate the whole problem. For example, with a head-up display, there is no decisionmaking problem of the kind we have been talking about because the head-up display provides a symbolic "real-world" display to the crew. In this way, the decisionmaking problem associated with transition from panel instruments to the outside world would probably be eliminated. There are many other aspects of the problem we could go into. We have been thinking (for example, in the category II problem) about subtle effects such as differences in the visual-performance characteristics of the two crewmembers, the captain and the copilot. I believe TWA handles category II problems by having the young copilot try to establish visual contact. He is looking outside. When he establishes visual contact the older, and perhaps myopic, captain looks up from the panel to the outside and he may not see the runway. It may turn out to be important to establish visual-performance compatibility between the two crewmembers in allocating task functions. The point I wanted to make is that the decision problem, though important, is not the whole story.

YNTEMA: That bears heavily on the second question I was asking, which I think is a crucial one in launching into a human-factors effort. Is it possible that this whole problem will vanish with the head-up display, that the problem of breakout and decision is going to be so modified?

RATHERT: I think that there is an easy straightforward answer. If you get the head-up display (let us postulate that a genius in the display group comes up with a perfect working head-up display and puts it in the cockpit), what will you do with it? The minimums will drop from 200 feet to 100 feet, because the airlines will then have a broader category in which to complete their scheduled flights. The airlines, through economic pressure, will say, "Fine, we have a head-up display. It helps a lot. We will move down to 100 feet." And you will have exactly the same decision process.

SADOFF: One of the potential problems with head-up displays is lack of registration. There might be the false-alarm kind of situation that Swets summarized and the decision would have to be made on how closely to register the head-up display to avoid "hypochondria" indications of system malfunction. For a category II flight situation, experts in the field like Naish at Douglas have established acceptable out-of-tolerance registration of the head-up display with respect to the real world. Additional tests are

required to establish whether a decisionmaking problem would exist with regard to the crew's ability to detect a real malfunction.

RATHERT: May I also respectfully point out that it will be a long time before you are going to have a head-up display in a Taylor Cub. Those people are just as precious to us as the airline pilots.

SWETS: Will you tell me what "category II" is?

SADOFF: "Category II" is the jargon that defines a certain reduced-visibility condition defined by a series of numbers: the altitude at which you break out (150 feet), and the slant range (the runway visual slant range when you first establish contact with the runway lights, about 1250 feet). In other words, that is the visibility you have: 1250 feet slant range from the aircraft to the runway, and 150 feet to see the ground if you look directly below. In "category III" there are several subcategories; but, in decreasing visibility, they run through *a*, *b*, and *c*. I do not know the specific numbers, but category IIIc is completely zero-zero. You may even need guidance to taxi back to the terminal.

RATHERT: If you are cleared for a category II landing, and as you come down to 150 feet you cannot establish visual contact with the runway, then you cannot proceed with the landing. You must go around. In other words, the categories are premade decisions that the conditions have to meet a certain minimum before a pilot can proceed with that type of landing. The total airplane system—and to a certain extent each pilot, through our rather complicated licensing system—is cleared for category II operation, or category III operation, and so on. The airports likewise.

SWETS: Have we established that the problem is real and will be here long enough for human-factors research to do something about it?

WARD EDWARDS: I worry about the identification of that particular problem as the one to put concerted effort into because it seems to me that, in the first place, it is, politically speaking, more than usually messy. In the second place, because it mixes what I see as many different varieties of human-factors problems into the same pot, it may have advantages; but, if we are talking about decision processes, such a mixture may have some disadvantages. Is there some reason why we are fixing on that problem?

SWETS: No.

YNTEMA: Belsley said several times that it was the problem.

BELSLEY: No. I said there were several kinds of problems. Let me make my position clear. I guess I have not. I am not necessarily saying that this group should work on this specific problem, but I would like to see the fruits of the basic research on decisionmaking applied in some way to a real-life problem. Edwards is engaged in trying to determine some coefficients, or to sharpen the focus, on his condi-

tional probabilities within a real-life situation. And, hopefully, when he gets these things, he will be able to apply it to a situation that is even more complex than the one he has been working with. Maybe he will not, but this is the intent. I have been hoping that since the airplanes are here to stay and are getting bigger and better and more complicated, that we can apply the fruits of decisionmaking research to these kinds of problems also. And for me to say that this problem is the only one it should be applied to is not necessarily true, but I would like to see someone take a crack at it. This is the one, as I see it—this breaking out and trying to decide whether thee shall go around with me or not at that point in time—that is very crucial. And a lot of people are flying behind the pilot. It is going to make a big splash some day, and maybe you will be flying behind the pilot. You certainly are not going to be flying behind an astronaut.

MARKOWITZ: About how many times a year is this decision made under category II conditions, the decision to go around?

EDWARDS: Go-arounds are very infrequent.

MARKOWITZ: I want to know how often that decision has to be made.

BELSLEY: He wants to know how many landings they make under category II conditions, that is all he is asking. They make a lot of them depending on the location. At London they are making them all the time.

YNTEMA: How many a day, probably?

BELSLEY: Well, you know what happens, what they do. Let me put it this way: For example, every 30 seconds they bring an airplane into Chicago O'Hare. The minute the weather gets sticky, the time required goes up to a minute and a half, and then the controllers start to stack airplanes. As the ceiling gets lower, the stacking gets worse. These things are not just unique, they occur all the time.

MARKOWITZ: On a bad day at O'Hare in 2 hours, you might say 100?

BELSLEY: You could bring that many aboard, but not necessarily.

YNTEMA: There are probably less than 1000 per day in the United States?

BELSLEY: But it ties up airports.

MARKOWITZ: The question is how much data could one expect to get in a real situation per unit time?

BELSLEY: I do not know that you are going to get it in a real situation.

RATHERT: Assuming you could use the actual landing under category II conditions, you would have at least 1000 per day. I say this with some confidence because they have set out to record landings, and this sort of thing, and this is the kind of data production you get out of it. Noise measurements are another example. The approach must be

made in a different way because of the noise problems. They record these.

SWETS: Let me clarify one thing. It was not in my mind nor Belsley's that the people who happen to be in this room at this time without exception focus in concert on a particular problem. On the other hand, we came here expecting to be somewhat frustrated, and indeed are, with respect to applications of research. The question is really what might a reasonable step be to let us feel that we were venting some of that frustration. It is not enough to talk about how poorly we are doing, or about how well we are doing if you would rather take that tack. A reasonable question is: What might we be doing?

YNTEMA: Before going to that, could I make a remark?

SWETS: We were about to leave that.

YNTEMA: I would think that, generalizing from other fields, the way that information crosses that boundary from research into advanced development is not very often by the push from the research man. One can always think of spectacular examples, but usually I think it is someone reaching from the other direction toward the research. I would suggest that it is someone who has been trained to think about where you can make a big difference in a system, as opposed to putting in a lot of effort and making a small difference. If you push real hard, you might make a 1-dB difference. The way to look for the practicality of the applications, I would suggest, is more in some direction of the sort that Birdsall has symbolized, arranging that the reach come from the other side. It is more a matter of people who know the problems reaching toward a class of research things than a research man saying, "Here is something I can apply," because he will not think of something like the head-up display. He will not think plies only to a corner of the problem. The creativity of the fact that his particular field of research ap- here, and it takes a lot of creativity to do this kind of thing, must come from the basis of knowledge of what the problem is. That is, the development man must be central, because usually this kind of practical thing is done by reaching for research results in a number of fields.

RATHERT: I am sure you will give us credit for pushing on both sides. We are pushing both sides equally hard. I think somebody used the word "frustration," too, which is exactly my state and the state of many people like me. Frustration at not being able to get these two together is so high that I am willing to push you toward the middle and push the development man toward the middle equally.

YNTEMA: I would suggest that you put your energy into pushing from the other side. No amount of urging research men to phrase their results in a form that will be accessible to people concerned about practical applications, or urging them to use their imagination and creativity to think of practical applications, is going to accomplish nearly as much as reaching from the other side.

R. MARK PATTON: I feel compelled to take the position that more will be gained by pushing the researchers. Maybe this is because it is what we do more often, simply because we deal with the researchers more than with the operational people. But we do not feel that we have to make unreasonable demands—we simply feel that we must keep the elbow in the back a little bit. When someone such as Edwards actually likes to have our elbow in his back, that is ideal from our point of view. You have to remember that NASA is a mission-oriented agency, and is not in the business of doing sky-blue research for its own sake. Some people think that I believe in supporting research with no thought of application, but this is simply not true. When anyone is supported by a mission-oriented agency, it seems to me that there is an obligation to consider that agency's needs. This means that one's job is not really done until the results are in a useful form. I might add that I encourage people to try for the application, without destroying the original format of their work, which I think is different from starting with the application, and tailoring the research to fit. I have no objection to the latter course, in fact a portion of our research is generated this way. But I feel that some of our best ideas may come from people who start in the laboratory, and attempt to apply their findings to some real-world problem. So I think that a portion of our program should be developed on this basis.

YNTEMA: Could I say that I think it has been a great meeting. I think it has followed very much the pattern that you were laying out. For someone not involved in this dialog, I find that this has been a very impressive interchange.

However, a tone seems to have crept in from time to time: "Would it not be nice if the people who are doing this research could have a brainstorm, a creative idea about how it could be used." What I am suggesting is that they just are not mentally equipped to do it. It is not that they do not have the brains, but they do not have, and practically cannot have, the information on which that creativity is based, except in a few odd cases, like someone who becomes a flight controller. Those cases would be rare.

PATTON: I just cannot see it. You know, if company A wants to hire a pilot to tell them about categories II and III, pilots can be hired. If university B wants, in the future, to turn in on their grant proposals an amount of money for consultation by an airline captain, this can be done. I think the sources of information are there. I do not think you must necessarily become a pilot to know about airplanes. I think the essence of fruitful dialog is in bringing different categories of information together

into one body. No one person can have every capability, so we must use a strategy to get what we need to solve a problem.

I think some general remarks just by way of thanks for your attendance are in order. As Tanner said, some of you in particular were not bound to come to us because you are not our grantees or contractors. I guess even the grantees and contractors could have said that they had the flu and not come. Anyway, we do thank you for attending. We thank you for your thoughts and ideas. My fear of a meeting such as this is not that we will have another one of a similar nature and make the same mistakes, but that it will not be exploited properly, because I think that is the essence of the problem of meetings. For 3 days everyone thinks great thoughts, has a great time, and then goes back to what he was doing in the first place. I am anxious to keep driving at the result in the sense of—exploiting is a poor word —in the sense of using the things that we have developed in the best way.

SWETS: The final comment; I want to add my thanks to Patton's, to the speakers, and also to thank our hosts. Thanks very much for 3 good days.

Participants

CONFERENCE CHAIRMEN

JOHN A. SWETS
JOSEPH MARKOWITZ
TRIEVE A. TANNER, JR.

DR. EARL A. ALLUISI
Department of Psychology
University of Louisville
Louisville, Ky.

DR. RICHARD C. ATKINSON
Institute for Mathematical Studies in the
Social Sciences
Stanford University
Stanford, Calif.

STEVEN E. BELSLEY
Deputy Asst. Dir. for Life Sciences
NASA Ames Research Center
Moffett Field, Calif.

DR. THEODORE G. BIRDSALL
University of Michigan
Ann Arbor, Mich.

GEORGE N. CHATHAM
Chief, Program Analysis Branch
Office of Advanced Research and Technology
NASA Headquarters
Washington, D.C.

DR. STANLEY DEUTSCH
Chief, Man-System Integration
Branch Office of Advanced Research and
Technology
NASA Headquarters
Washington, D.C.

DR. WARD EDWARDS
Engineering Psychology Laboratory
University of Michigan
Ann Arbor, Mich.

DR. JEROME I. ELKIND
Bolt, Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Mass.

DR. EDWARD M. HUFF
Human Performance Branch
NASA Ames Research Center
Moffett Field, Calif.

DR. LLOYD A. JEFFRESS
Defense Research Laboratory
University of Texas
Austin, Tex.

DR. ALFRED B. KRISTOFFERSON
Department of Psychology
McMaster University
Hamilton, Ontario

DR. JOSEPH MARKOWITZ
Bolt, Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Mass.

HAROLD G. MILLER
Chief, Mission Simulation Branch
NASA Manned Spacecraft Center
Houston, Tex.

DR. R. MARK PATTON
Chief, Human Performance Branch
NASA Ames Research Center
Moffett Field, Calif.

GEORGE A. RATHERT
Chief, Simulation Sciences Div.
NASA Ames Research Center
Moffett Field, Calif.

JOHN W. SENDERS
Bolt, Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Mass.

DR. JOHN A. SWETS
Bolt, Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Mass.

DR. TRIEVE A. TANNER, JR.
Human Performance Branch
NASA Ames Research Center
Moffett Field, Calif.

DR. DOUWE B. YNTEMA
Lincoln Laboratory
Massachusetts Institute of Technology
Lexington, Mass.

FIRST CLASS MAIL

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

— NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C. 20546