HIGHWAY RESEARCH RECORD

Number 55

Driving Simulation 1963

6 Reports

Presented at the
43rd ANNUAL MEETING
January 13-17, 1964

HIGHWAY RESEARCH BOARD
of the
Division of Engineering and Industrial Research
National Academy of Sciences—
National Research Council
Washington, D. C.
1964

Special Committee on Driving Simulation

(As of December 31, 1963)

Carlton C. Robinson, Chairman
Director, Traffic Engineering Division
Automotive Safety Foundation, Washington, D. C.

Burton W. Stephens, Secretary
Research Psychologist, U. S. Bureau of Public Roads
Washington, D. C.

Kenneth D. Chacey, Director of Transportation Engineering, Office of the Chief of Transportation, Department of the Army, Washington, D. C.
Ronald L. Ernst, Research Associate, Engineering Experimental Station, Ohio State University, Columbus
Theodore W. Forbes, Assistant Director (Research), Highway Traffic Safety Center, Michigan State University, East Lansing
Bernard H. Fox, Consultant, Research Branch, Division of Accident Prevention, U. S. Public Health Service, Washington, D. C.
A. L. Haynes, Director, Advanced Product Study Engineering, Ford Motor Company, Dearborn, Michigan
Slade Hulbert, Associate Research Psychologist, Institute of Transportation and Traffic Engineering, University of California, Los Angeles
John C. Kerrick, Director, Drivers' License Program, American Association of Motor Vehicle Administrators, Washington, D. C.
Barry G. King, Chief, Research Branch, Division of Accident Prevention, U. S. Public Health Service, Washington, D. C.
James L. Malfetti, Executive Officer, Safety Education and Research Project, Teachers College, Columbia University, New York, New York
Burton W. Marsh, Director, Traffic Engineering and Safety Department, American Automobile Association, Washington, D. C.
Theodore F. Morf, Assistant Chief State Highway Engineer, Illinois Division of Highways, Springfield
Henry M. O'Bryan, The Bendix Corporation, Washington, D. C.
E. S. Preston, E. S. Preston & Associates, Washington, D. C.
C. W. Prisk, Deputy Director, Office of Highway Safety, U. S. Bureau of Public Roads, Washington, D. C.
David W. Schoppert, Associate, Alan M. Voorhees and Associates, Washington, D. C.
J. C. Womack, State Highway Engineer, California Division of Highways, Sacramento

Contents

Time Sharing Between Compensatory Tracking And Search-and-Recognition Tasks

B. W. STEPHENS and R. M. MICHAELS
Human Factors Research Branch, Traffic Systems Research Division, Office of Research and Development, U. S. Bureau of Public Roads

This study deals with one aspect of driving behavior; i.e., the sharing of time between two types of activities that the driver is forced to accept as a part of the driving task. Steering and recognition behavior has been abstracted and explored in the laboratory by analyzing performance on a compensatory tracking task taken as an analogue of steering and on a filmed presentation of a sign search-and-recognition task. The relative effects of the interaction between tasks were explored. The main findings of the study were that (a) where time sharing was required, each type of performance was degraded; (b) increasing the number of message units appearing in the recognition task did not differentially affect simulated steering performance but did increase the time required for recognition of a key message; (c) increased speed of the simulated steering task displayed decreased recognition time of the discrete visual task; and (d) where a specific message had equal likelihood of appearing or not appearing, recognition time was greater when the key word did not appear. Results are discussed in terms of operator sampling behavior.

•CONTROL of lateral placement on the roadway is an intrinsic part of the task of driving which requires the human operator to maintain vehicular position within rather narrow bounds. In addition, the driver must search for, select, and process a variety of discrete informational indicators relevant to his ongoing and future psycho-motor and decision-making performances. These two tasks require two very different visual tasks which compete for the viewer's time and must in some way be sampled such that minimal variation in steering is achieved. How these two types of operator performance interact and the effects of their interaction on system performance are poorly understood. Yet, the efficiency and safety of the system depend on the proficiency of human processing of these concurrent informational inputs.

The experiment described here provides data relevant to the operator's sharing of time between a continuous control task, referred to as tracking, and the task of seeking out and detecting a critical aspect of a rather heterogeneous environment, referred to as search-and-recognition activity.

Previous investigators have considered each of these two aspects of operator performance, but have not examined the interaction processes. A more formal definition of both tracking and search-and-recognition behaviors, along with studies relevant to this inquiry, provides some basis for predicting the baseline conditions employed in this research.

One-Dimensional Tracking

Tracking has been assumed to be an analogue of the steering task and is therefore defined in this context (1); that is, as a paced task where an "externally driven input

Paper sponsored by Special Committee on Driving Simulation.

signal defines an index of desired performance and the operator activates the control system to maintain alignment of the output signal of the control system with the input signal. The discrepancy between the two signals is the error and the operator responds to null the error."

Compensatory tracking provides a signal of the error and requires only a reference and control mechanism. Such a situation is analogous to minimizing deviations from a prescribed path or roadway. Such a configuration may be displayed on the face of an oscilloscope. Use of a low frequency sinusoidal input provides a geometrically smooth input similar to roadway curvature. The information load appears to be a function of the increment magnitude that must be achieved within a relatively stable response period. As the frequency increases, the information load increases. Consequently, any time consumed in the evaluation of other portions of the environment would be expected to lead to more imperfect proficiency of guidance.

Subjects appear to be able to match input frequencies up to 3 or 4 cycles per second (cps) according to Ellson and Gray (10). McRuer and Krendel (14) give a value of about 1 cps. A specific control or display configuration may affect that motor response sufficiently to account for these differences.

In a field situation, amplitude has been shown to affect the operator's error-free performance of guidance. A study at the Ford Motor Company (13) indicated that on a test track with sine wave configurations, drivers reduced their speed to effectively track higher amplitude curves. This field study suggests that increments of angular change remain constant over a time span probably associated with the subject's response time.

Brown (7) has shown that many cases involving objects in motion may be considered to be perceived as a constant proportion of the velocity of angular change, suggesting that a single metric may suffice to predict tracking error for a particular control configuration.

Recognition Time

Recognition has been defined by Munn (15) as the ability "to differentiate between the familiar and the unfamiliar or what has been experienced and what is new." Recognition time may be defined as the delay between the onset of some information display and its overt discrimination by an observer. In the context of this research, search-and-recognition activities were measured. The distinction between simple recognition and the dynamic task discussed here is that the task is not localized. It is in motion and must be pursued at least briefly. It must indeed be searched for, pursued, recognized and finally reported by some overt act.

In the static state, discriminations of this type have been measured by obtaining measures of disjunctive reaction time. Considerable evidence supports the assumption that recognition time increases with task complexity. Classical reaction time experiments indicate that reaction time increases with the number of alternative responses (9, 16). In an unpublished study, performed at the U. S. Bureau of Public Roads, Desrosiers (8) found in a simulated situation that the mean recognition distance of a subject approaching a sign at a constant velocity varied inversely with the number of alternative messages on highway signs.

Task Sharing

Many studies have been conducted where information from separate discrete tasks is to be optimally organized and translated to control devices by the human operator (6). Most of these have involved the sharing of tasks between more than one sensory modality.

Briggs and Howell (5) have provided data on the interaction of two continuous tracking activities considering the effects of stimuli separation and speed. They have considered both peripheral and central aspects of the time-sharing process. Although more than one information source was to be attended during this study, three major differences existed between their study and the one reported here. First, both tasks required tracking perfomance which demanded forced pacing in an alternation pattern between the inputs by the operator unless he left unattended one of the two components of the time-sharing situation. Second, the spatial separation of displays was fixed for any one experimental condition, reducing search time and also the applicability of such a configuration to the situation in question; i.e., utilization of information sources which are themselves in motion, relative to the observer. Third, neither task required the discrete decoding of visual information implicit in recognition tasks. There are, however, certain requirements met by the study by Briggs and Howell which have pertinence to the situation of prime interest. These are discussed at a more appropriate point in this report. All studies reviewed indicated a generalized degradation of performance as task complexity increases.

The following statements concerning time-sharing performances formally present the major hypotheses subjected to test in this study.

1. The effect of a superimposed search task will increase the tracking error (degrade tracking performance) in proportion to the frequency and amplitude of the tracking signal (a sinusoidal function).

a. Tracking error will increase proportionally with the tracking frequency and amplitude.

b. Tracking error will increase proportionally with the number of words on the signs (complexity of the search-and-recognition task).

2. The speed and accuracy of the search-and-recognition task will be inversely proportional to the frequency and amplitude of the tracking task and to the increase in the complexity of search-and-recognition task.

a. Recognition time (measured as time from the sequence beginning) will be proportional to the frequency and amplitude of the tracking signal (a sine wave).

b. The proportion of false alarms in recognition will increase as the frequency and amplitude of the tracking signal are increased.

3. Tracking performance will be decremented equally when a specific message in the sign display is present or not present for signs bearing equal numbers of alternatives.

a. Tracking error will not be dependent on the presence of a specific message on the sign.

These hypotheses suggest monotonic changes and do not predict absolute magnitudes of error. They are intended to ascertain the relative effects of the concurrent operations that subjects were requested to perform.

Figure 1 presents the three-dimensonal matrix of conditions defining the range of variables composing the experiment. The matrix does not differentiate between experimental conditions providing the presence or absence of a specified message.

The separation of experimental and control conditions is also shown in Figure 1. Quantative predictions of tracking error are suggested by the frequency-amplitude product multiplied by a constant. No quantative predictions were made for search-and-recognition times.

The first hypothesis predicts increasing erroneous tracking performance related to an increasing frequency-amplitude product. The lower left-hand corner of the matrix reveals the lowest such product, hence predicts the smallest tracking error. The upper right-hand block displays the greatest tracking error. This trend is paralleled for each of the search-and-recognition conditions with a corresponding increase in the predicted tracking error as the number of words per sign increase.

The second hypothesis predicts a positive relationship between recognition time, as a measure of search-and-recognition performance, and the frequency-amplitude product of the tracking task. Moving from the upper right-hand corner diagonally across to the lower left-hand corner of the matrix of conditions is a prediction of increasing recognition time for signed key messages, which is related to frequency and amplitude across conditions of increased number of words per sign.

The last major hypothesis tested in the study relates to the influence of expectancy of appearance of a message of the search-and-recognition and its relation to tracking between the presence or absence of a key message.

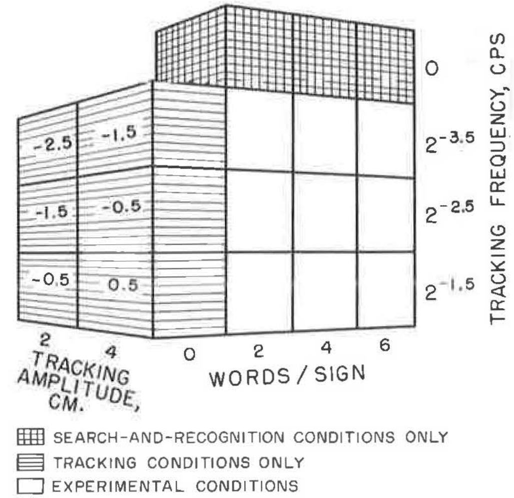


Figure 1. Matrix of experimental conditions.

METHOD

Apparatus

Figure 2 shows the general relations of the task configuration to the response devices. Each of these blocks in their interacting role is discussed in this section. In addition to the presentation and response systems, timing, measurement and recording systems for the experiment are discussed. Figure 3 shows the relative position of the displays and control devices.

Tracking Task

The subject was seated at a console containing a circular cathode ray display (CRT), 12.5 cm in diameter. The display was situated slightly below the line of sight at a 12° tilt with the vertical axis of the room. A low-intensity ambient brightness of approximately 10^{-3} mL surrounded the scope face. The brightness of the tracking signal was approximately 1 mL.

The signal was a point of light normally driven horizontally across the face of the CRT and was to be corrected by manipulation of a simple manual rotary control (a handcrank). The control mechanism, 15 cm in diameter, had a 90-kg flywheel attached to prevent jerking of the control system. Two such cranks were located on the console so that either one could be easily operated by the subject's preferred hand. The two controls were mounted on the face of the console at a 45° angle with respect to the floor. The crank was adjustable to a height which required support for the elbow by a rubber pad. Hence, rotary control was achieved almost entirely by wrist motion.

Search-and-Recognition Task

A series of motion pictures containing scenes of a new section of the Interstate Highway near Washington was presented on a rear-projection ground glass screen. The

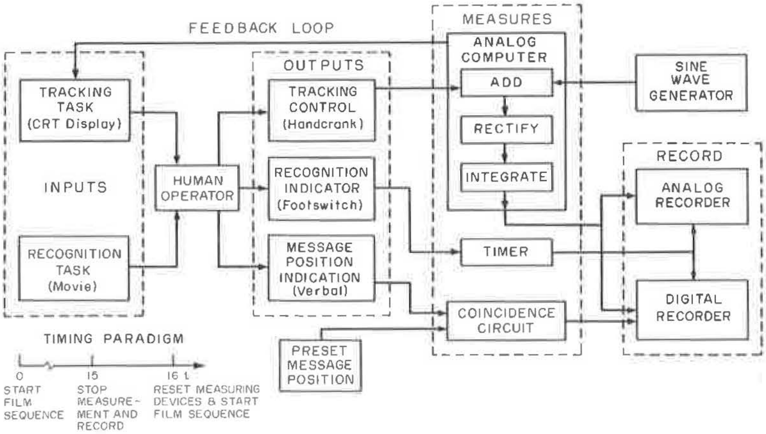


Figure 2. Block diagram of time-sharing system.

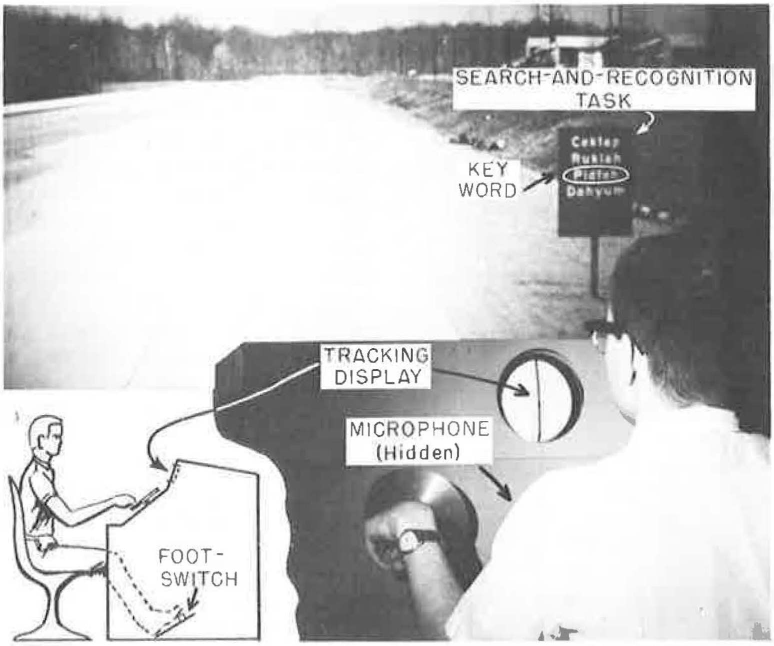


Figure 3. Subject seated at driving console performing tracking and search-and-recognition tasks.

screen was located 8 ft from the seated subject with vertical alignment to the CRT. No apparent texture masked the projected images and the angle of inclination required for a visual shift from one display to another was slight.

Displayed on the face of the screen was a sign-reading task containing messages composed of nonsense syllables having association values of 27 to 47 percent. Each message unit had six letters and the lengths of the messages and the contours of the letters composing the message units were matched.

The apparent sizes of the signs were the same as they would have been on the actual highway configuration; i.e., the visual angle subtended by the sign on the operator's eye was matched for the time from the sign that corresponded to the distance of an actual vehicle traveling at constant velocity. Each approach scene was projected at a rate of 24 frames per sec, and was 16 sec in duration including 1 sec of film displaying passing and receding from the sign position.

Before photographing the highway section, portable sign standards were erected with green backgrounds nearly reproducing the hue and saturation of the present standards used for destination signs on the Interstate Highway System. During photographic sessions, interchangeable messages were placed on the standard backgrounds and the signs were oriented so that the brightness was approximately equal for all signs. The camera was mounted near the head position of the driver, hence providing motion pictures which were used to synthesize the appearance of driving on a roadway from the driver's view.

During photography, four standards were placed 1,000 ft apart with one of 30 sign configurations randomly assigned to each location so that background environment could not be associated with a particular sign configuration during the experiment. Each sign contained a combination of 2, 4 or 6 message units with or without the presence of a key message, PIDFOH, and position of the key message was also randomly assigned to one of the available positions.

The developed pictures were duplicated so that seven complete randomized edited versions were available for the experiment. All films had the 30 scenes of the randomized standardized signs in different randomized orderings.

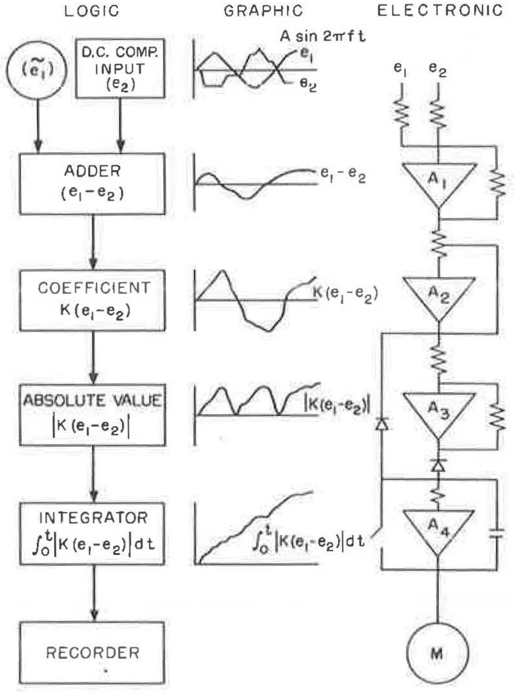


Figure 4. Method for recording tracking error, represented in three ways.

Timing

The timing system provided sampling of performance on the two time-shared tasks. The timing paradigm is shown in Figure 2. Synchronization of time was achieved by employing the film as a common time base. A small hole was cut in the corner of frames spaced in time 15 and 1 sec apart. Light from the projector source energized a photo-resistor in the upper corner of the screen. The subject seated on the other side of the ground glass screen could not see flashes of light from the projector and hence perceived a synthesis of a continuous roadway with no interruption.

The photo-resistor pulsed a "flip-flop" which alternately: (a) reset the measurement devices to a de-energized state, followed by an activated state of measurement, and (b) stopped the measurements and digitally recorded performances on each of the subject tasks.

Measurement

Measurements of rectified integrated tracking error, time from the sequence start where subjects indicated recognition associated with the sign reading task, and correctness of verbal responding to the key word on the sign reading task were recorded digitally. These measures were employed as the dependent variables for determining the tenability of the time-sharing hypotheses. Analogue records of tracking performance and a discrete indication of recognition time were also collected.

The system of measurement of cumulative tracking error is shown in Figure 4. This error measurement system is essentially the one developed by Gain and Fitts (11) for a series of tracking studies. In this system, one of the cranks was attached to a linear potentiometer which controlled a d.c. voltage. This in turn was added to a low-frequency sine wave whose algebraic sum was displayed on the cathode-ray display and in parallel course multiplied by a constant value empirically derived for minimal error of computation, rectified and integrated. An output pulse from the flip-flop actuated a double-pole, double-throw relay between the rectifier output and the integrator at the termination of each 15-sec sampled sequence. This relay in turn energized a delay sample circuit associated with a Hewlett-Packard Model 405-AR d.c.-digital voltmeter. The voltmeter in turn "issued a print command" to a Hewlett-Packard Model 560-A digital recorder. Total tracking error system measurement during any one day of running was less than 1 percent.

The sampled sequence initiation pulse from the flip-flop actuated another double-pole, double-throw relay which issued a short pulse to the clamping circuit associated with the integrator, thereby "zeroing" the unit and permitting an accurate accumulated error score to be developed.

The measure of time from the end of the sampled sequence was achieved by use of the second set of contacts on this pulsed relay. A positive going pulse of 80 v started the count of a Hewlett-Packard Model 522-B electronic timer. A stop pulse was initiated by either the subject's application of pressure on a foot switch or by the one set of relay contacts associated with the stop phase of the flip-flop.

Verbal indication of the position of the key word on the search-and-recognition display was measured for "correctness" of response by use of a coincidence circuit. For each sequence the experimenter set into a series circuit the position associated with the key word at the beginning of each trial. The verbal response was recorded by pressing the second switch in the pair of cascaded switches. If the position reported was the same as that pre-set by the experimenter, a voltage activated a binary indicator and the digital recorder printed an indication of position correctness.

Procedure

During the experimental runs, subjects were requested to simultaneously track (keep the dot on the CRT display centered on a vertical black line dividing the scope face) and to indicate when the word PIDFOH could be recognized on the motion picture display. Subjects pressed the foot switch as soon as they recognized PIDFOH, or if the key word was not recognized as present, they were also to press the switch. Immediately upon initiating the motor response to the search-and-recognition, subjects verbally gave the position of the key word by reporting the numerical position, one through six, or "zero" when PIDFOH did not appear on the sign.

Subjects were not instructed to pace the task in any particular manner, but merely to give indications of their recognition of the key word as soon as possible without guessing and to track as well as possible throughout each trial.

Prior to the experimental session, subjects had approximately 4 hr of training in tracking and search-and-recognition. The film used for the recognition control cases was also used for training. In all cases, with the exception of the last subject, training took place the day preceding the experimental session. The experimental session was divided into two 3-hr sessions. Each of the 14 trials was 8 min in duration with rest periods of approximately that length between trials.

Six subjects randomly drawn from the Bureau of Public Roads Office of Research Staff were employed. The four males and two females, from 21 to 45 years old, reported no anomalies of vision or motor impairment and all had at least several years of driving experience.

Experimental Design

A balanced five-factorial design, with randomized presentations of the stimulus conditions associated with the search-and-recognition task, was randomly presented for each tracking condition. Six subjects were employed for tracking under two amplitude and three frequency conditions under filmed conditions having signs containing 2, 4, or 6 messages with or without the presence of a key message PIDFOH. Each of these conditions was presented five times. The control conditions associated with tracking were equal in length, but the sample size was accordingly increased to 30 trials.

A latin square treatment of frequency, amplitude and subjects provided a determination of time transitions that might be associated with the experimental presentations.

Control tests of tracking performance were paired with each of the time-shared (experimental) runs. The first half of the trials provided a control case requiring tracking without the filmed presentations followed by an experimental case where subjects were required to perform both tracking and search-and-recognition activities. The latter half of the trials for each subject was a reversal; i.e., each experimental trial was followed by a control case.

Two control trials providing recognition information only were conducted during the first and second halves of the experimental session. No tracking was required.

RESULTS

Four types of analysis were conducted to ascertain the influence of constituents of the tracking task on performance on the search-and-recognition task and likewise the influence of components of the search-and-recognition task on tracking performance. These analyses included testing of the significance of the differences between means of each of the treatments described earlier; i.e., an analysis of the integrated tracking error variances and similar analysis of the recognition times between each of the conditions associated with the dynamic recognition task.

Basic to the inquiry was conduct of a test of the generalized time-sharing hypothesis which states that there is a trade-off between the two tasks such that high erroneous performance on one of the tasks is highly correlated with relatively accurate performance on the other. This hypothesis was simply tested by sampled rho-correlations of the time-shared aspects of the experimental configuration.

A third type of analysis of time-sharing behavior of the interpolated tasks was an examination of the erroneous recognitions of signs presented. Finally, control tests were conducted to ascertain: (a) possible learning confounding throughout the experimental sessions, and (b) a comparison of experimental and baseline conditions where the tracking task or search-and-recognition task was presented singly.

The tracking data and recognition time data are separately presented and finally the interactions of the two performance modes are considered together in an examination of the "trade-off" hypothesis.

Tracking Performance

A summary of the factors and their interactions interjected into the experimental situation is given in Table 1. This table does not reflect analysis of presentation of the tracking task alone, but rather those presentations where subjects were required to perform both tracking and search-and-recognition tasks concurrently. Inferences therefore drawn from these data reflect not only the effects of the attribute in question but also the influence of attributes of the search-and-recognition task. Table 1 shows that frequency, amplitude,

TABLE 1

ANALYSIS OF VARIANCE SUMMARY TABLE FOR CUMULATIVE TRACKING ERROR

Factor	Sum of Squares	DF	Mean Sum of Squares	F Ratio
Frequency	5, 582.52	2	2, 791.26	493.15[a]
Amplitude	4, 593.27	1	4, 593.27	811.53[a]
Complexity	22.58	2	11.29	1.99
Presence	59.52	1	59.52	10.52[a]
Subject	2, 481.67	5	496.33	87.69[a]
FxA	2, 244.01	2	1, 122.00	198.23[a]
FxC	33.85	4	8.46	1.49
FxP	14.01	2	7.00	1.24
FxS	490.69	10	49.07	8.67[a]
AxC	1.06	2	0.53	0.09
AxP	12.97	1	12.97	2.29
AxS	635.52	5	127.10	22.46[a]
CxP	26.58	2	13.29	2.35
CxS	35.39	10	3.54	0.62
PxS	19.82	5	3.96	0.70
FxAxC	22.10	4	5.52	0.98
FxAxP	13.69	2	6.84	1.21
FxAxS	509.41	10	50.94	9.00[a]
FxCxP	60.59	4	15.15	2.68[a]
FxCxS	124.93	20	6.25	1.10
FxPxS	50.21	10	5.02	0.87
AxCxP	12.76	2	6.38	1.13
AxCxS	49.89	10	4.99	0.88
AxPxS	25.47	5	5.09	0.89
CxPxS	87.22	10	8.72	1.54
FxAxCxP	54.24	4	13.56	2.40
FxAxCxS	108.15	20	5.41	0.96
FxAxPxS	60.65	10	6.06	1.07
FxCxPxS	146.28	20	7.33	1.30
AxCxPxS	17.23	10	1.72	0.30
FxAxCxPxS	157.80	20	7.89	1.39
Residual	4, 894.44	864	5.66	-
Total	22, 676.20	1, 079	-	-

[a]Significant with probability less than 0.01.

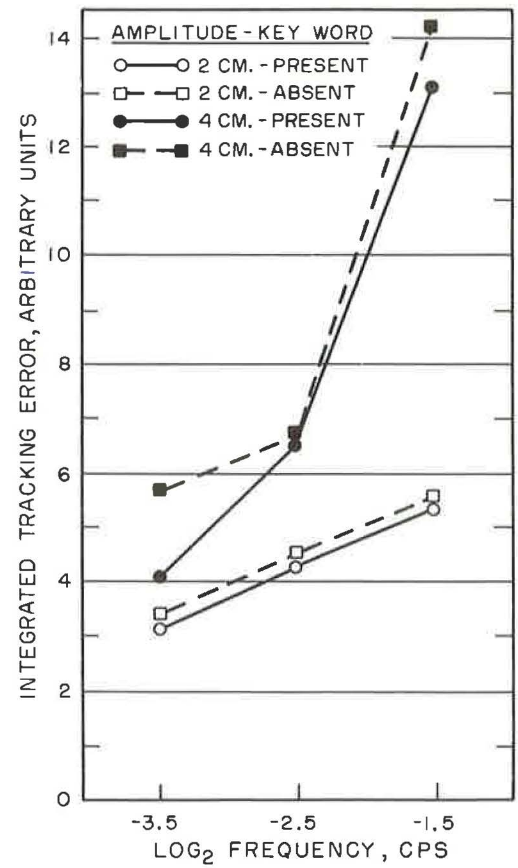


Figure 5. Relative tracking error data showing effects of frequency, amplitude and presence of search-and-recognition key word.

presence of the key word, and subjects have demonstrated differential effects greater than can be expected by chance one time in a hundred. Further, only the interactions involving frequency or amplitude and subjects and their respective and combined interactions display significance.

Figure 5 shows tracking error as a function of the significant variables of frequency, amplitude and presence of the key word. The data for all subjects are combined.

The differential effects of tracking error compared to the control condition (tracking without the time-shared task) are clearly demonstrated in Figure 6. This figure demonstrates the relative magnitude of differences and illustrates the presence of degradation due to the compounding of the two imposed tasks. There is a clear decrement in proficiency when the two tasks are facing subjects; in both the low and high amplitude cases. The error score in the dual task increases overall about 80 percent, whereas the differences between the key word's presence and absence are separated by only a small, but reliable amount (about 15 percent).

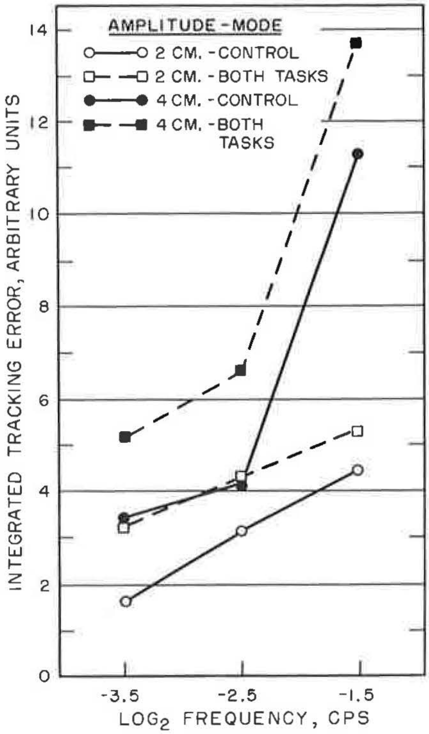


Figure 6. Relative tracking error data showing effects of sign reading task by frequency and amplitude of tracking signal.

Using the method of matched groups for all subjects, both presence and absence conditions of the key word and all sign complexities, a comparison by means of student's "t" test of the lowest frequency-amplitude condition with and without the interpolated task yields t = 3.92 which is significant at the 0.01 level. Therefore, tracking performance for the lowest average velocity condition without the secondary is superior to the comparable time-shared condition.

Search-and-Recognition Time

As observed for measures of tracking performance, subject differences for recognition time are also significant. Table 2 summarizes recognition time cases where the key message on sign was present.

Of the two components of the tracking task, frequency appears to have differentially affected performance on the dynamic recognition task. Figure 7 shows the time from the sequence beginning when subjects indicated recognition of the key message for each of the frequencies employed in the experiment. The reader is cautioned that this curve is illustrative because subject differences influence such a plot. It is shown that employment of tracking as a secondary task greatly increases the time required for recognition of the signs' key message.

The lack of subject interactions of complexity indicates that the subjects responded in essentially the same manner to different numbers of message units on the signs. The differences between two, four and six words reflect a monotonic increase in recognition time as the number of message units increase. The time increase between the two- and six-message signs is only about 4.4 percent or slightly less than ½ sec. The range of values for the time-sharing case extends from approximately ½ sec to about 2 sec beyond the recognition time for the control case where the film was shown without the displayed tracking task.

TABLE 2

ANALYSIS OF VARIANCE SUMMARY TABLE FOR RECOGNITION TIME

Factor	Sum of Squares	DF	Mean Sum of Squares	F Ratio
Frequency	9.53	2	4.76	4.81[a]
Amplitude	3.82	1	3.82	3.86
Complexity	20.61	2	10.30	10.40[a]
Subject	1,062.65	5	212.53	214.68[a]
FxA	5.84	2	2.92	2.95
FxC	2.29	4	0.57	0.58
FxS	28.22	10	2.82	2.85[a]
AxC	0.11	2	0.06	0.06
AxS	5.45	5	1.09	1.10
CxS	17.14	10	1.71	1.73
FxAxC	4.80	4	1.20	1.21
FxAxS	28.12	10	2.81	2.84[a]
FxCxS	27.48	20	1.37	1.38
AxCxS	11.13	10	1.11	1.12
FxAxCxS	28.89	20	1.44	1.45
Between treatments	1,256.08	107	-	-
Error	428.47	432	0.99	-
Total	1,684.55	539	-	-

[a]Significant with probability less than 0.01.

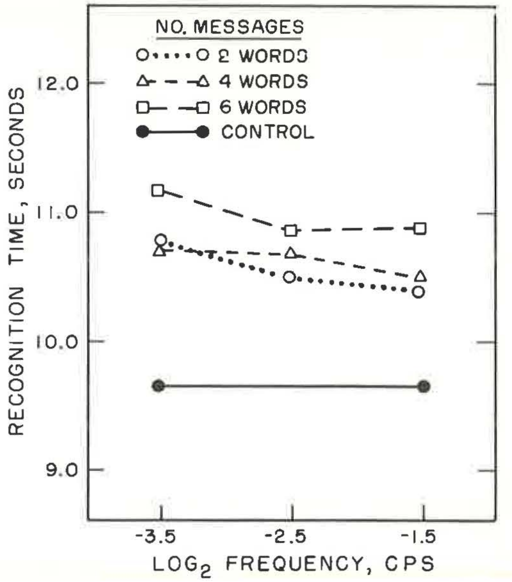


Figure 7. Recognition time data showing effects of time sharing with tracking task by frequency for different numbers of words on sign (compared with average time without tracking).

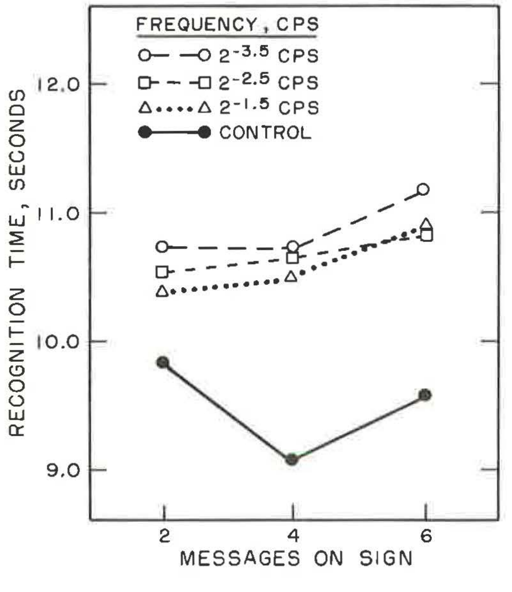


Figure 8. Recognition time data showing effects of number of messages and tracking frequency (compared with non-time-shared performance).

TABLE 3

PERCENTAGE OF ERRONEOUS RESPONSES BY SUBJECT

Subject	Error	False Alarm	Total Erroneous Response
1	0.8	1.7	1.2
2	5.8	2.5	4.2
3	2.5	0.8	1.7
4	1.7	0.8	1.2
5	0.8	0.0	0.4
6	0.8	0.8	0.8
Avg.	2.1	1.1	1.6

The control and experimental means are both shown in Figure 8. The control conditions show a dip at 4 message units per sign which does not parallel the general trend of the experimental data. The time differences are significant between 2- and 4-message units at the 0.05 level, but fail to reach significance between 4 and 6 units.

Search-and-recognition performance was tested to ascertain whether there was any change in performance over the first and second halves of the experimental sessions. Using a t-test for matched groups, the mean recognition times for each subject were compared for the 5th and 10th trials. No statistical differences at the 0.05 level were obtained.

The errors in judgment were obtained from the responses of the subjects to the position of the key word on the signs. The error scores are given both by percentage of misses and false alarms in Table 3. The table gives the error rates incurred by each of the subjects with respect to reports of erroneous position indications and reports of position when in fact the key message was not present (false alarms). These rates are sufficiently low that no test to discriminate between guessing and non-guessing was required. The total error was well below the 5 percent criterion imposed, therefore no additional analysis of the erroneous responses was conducted.

Time-Sharing Hypothesis

It was hypothesized that there should be a systematic inverse relationship between tracking error and recognition time if time sharing is determined on an individual trial basis. That is, if subjects optimize their performance on one of the two imposed tasks they would be expected to yield performances which were highly negatively correlated with each other. Randomly selected samples of performance were taken for time-sharing conditions for individual subjects. Twenty such samples were subjected to rank-order rho-correlations. Only one of the correlations reaches significance at the 0.05 level of significance. Therefore, rejection of the hypothesis that there is a significance correlation between the two time-shared performances must be made as the number of significant correlations does not exceed those expected by chance alone.

DISCUSSION

Time Sharing

The analysis of data revealed no systematic tendency on the part of operators to optimize one task to the exclusion of the other, when both tasks were simultaneously presented. Subjects were instructed to perform both tasks as well as possible and not to guess on the recognition task. Otherwise the experimental situation was relatively unstructured.

The analysis of time-shared process may, therefore, be derived from an analysis of the pattern of response and eye movements or from the magnitude and dispersion of errors of tracking and search-and-recognition performance. The latter alternative is within the domain of this study.

No effects of the number of messages, presented on the signs, on tracking performance are evidenced from the data. This would suggest that the overall time required by the dynamic recognition task was not affected differentially with regard to the number of message units. Yet the number of message units did contribute to increasing recognition time. Such a paradox either suggests that: (a) the differences in recognition, although significantly different are so small that they merely do not affect tracking performance (the variance associated with other factors is relatively large), or (b) the subject must quicken his alteration response to or from the tracking task to decrease time occupied in non-tracking activity.

Although differences in the differential error between time shared and control cases involving tracking alone are not great for each of the frequency-amplitude combinations, the trend is toward lesser differentials rather than greater ones as the tracking frequency is raised. If a constant differential between each control and experimental comparison is assumed, then it can be seen that the same increment of error is attributable to each of the frequencies.

With no difference between the integration of sine waves of differing frequencies (unless the integration time is very short) exists, the lack of visual feedback for 1 or 2 sec becomes increasingly important as the frequency is increased. Gottsdanker (12) has indicated that although essentially the same mechanisms operate in extrapolation of a course under visual and visually deprived conditions that operators manifest more variable and less accurate performance when they are visually deprived. The subject does maintain some proprioceptive feedback and is able to extrapolate the course, but this ability is decremented as frequency is increased.

Implications of Briggs and Howell's work (5) are that where there is spatial separation of tasks to be performed that a discrete sampling process must occur. Where disjunctive tasks compete for the observers' efforts, the sampling rate is paced in accord to the objective velocity of the tasks as well as some subjective criterion. A decrement in performance supposes a lack of visual error feedback, or expressed differently, a reduction of the sampling rate of the error signal on the part of the operator. If there is a rate of increase less than that expected by a condition of reduced feedback information, then the operator must increase his sampling rate with an associated effect on some other aspect of behavior. This assumption may be made only if the operator is presumed to be operating below the limits of his channel capacity.

It is therefore expected that if an equal time duration is expended on the search-and-recognition independent of the tracking task, that: (a) the analysis of variance of recognition times will show no significant differences for tracking frequency, amplitude, or interactions of those components (i.e., recognition times would not be increased or decreased due to the addition of the secondary task; but, as was seen earlier, the average recognition durations decrease as the frequency of tracking task increases); and (b) if an increment in tracking error in the time-sharing condition exists, then such an increment should increase as a function of increased frequency.

If the converse of both of these conclusions is shown, then it may be presumed that a time reduction on recognition is utilized in minimizing tracking error. The data clearly substantiate such a notion. The number of messages displayed to subjects shows no effects on tracking performance, but indeed frequency and the complex interaction of frequency-amplitude-subject are shown to have effects on recognition time at a level substantially beyond that expected by chance. A reasonable inference to be drawn is that as frequency (a component of angular change of the target) is increased, the rate of sampling of that course is increased with a consequent alteration in search-and-recognition behavior. The analogue to the field situation suggests that if the steering task may be considered in the same context as tracking, then search-and-recognition behavior will suffer with respect to sign reading complexity and actually improve as tracking velocity is increased.

The effects of the key word not appearing on the sign on tracking performance suggest that expectancy played some part in diverting the "attention" of subjects from the simulated steering activity. The significant increase in tracking error leads to the conclusion that greater periods of time were employed in attempting to find the key word in its absence. The implications to signing practice are discussed briefly in a later section.

Cues in Non-Time-Shared Recognition Trials

Noted in the comparison of control and experimental data was a decrease in recognition time associated with increased numbers of messages in the search-and-recognition where no time sharing is required. It would be predicted on the basis of Desrosiers' (8) laboratory results that an increase in recognition time would be required. Although Desrosiers found no significant differences in field tests, this lack of significance was attributed to the high variance of performance in the field situation. It is most likely that due to the use of the control film for training that disproportionate learning was associated with the four-message signs. There was no way of directly testing this hypothesis. However, close observation of these films revealed that activity was greatest in the periphery for the scenes where four-message signs appeared. Where subjects were required only to view the film (which in this case had been repeatedly shown during training), it is likely that such peripheral cues would be detected and utilized.

Spatial Separation of Information

Although care was taken to present a reasonable projection of the roadway (i.e., the rates of change of objects in the simulated view of the road were comparable to those in the field), fidelity of the films employed required much larger angles of message sizes subtended, for recognition. This, in effect, also meant that the lateral movement of objects across the viewed field was much farther from the line of regard in the laboratory situation than would be encountered in the driving world. Concurrently, the angular velocity of the signs was much higher at the time of recognition. This constraint meant that the message, when detected, was displayed from the line of regard about 10 deg; whereas in the field, previous tests under comparable speed conditions showed recognition of signs of the type employed in this study to be recognizable at about 400 ft, making the displacement of the sign at the time of recognition about 2 deg, within the foveal region of the eye. The shift of fixation in the field situation is quite small, thereby reducing the transition time between the two information sources, visual steering feedback and the messages on the signs.

Automatization

Control devices employed for repetitive inputs permit some reliable kinesthetic cues for the operator. The roadway has a wide range of cues which may be translated through the steering control, but such inputs appear to have low redundancy. In the experiment described there was high redundancy in the compensatory tracking signal; hence the error, incurred during periods when the subject directed his fixation toward the signs, was probably minimized. Bahrick and Shelly (3) found in tasks dependent on both visual and kinesthetic cues, that during prolonged trials a gradual change from extroceptive cues to proprioceptive control occurred. So indeed there may have been a reduction in the error scores under these high redundancy conditions. A balanced design reduced the influence of this factor in analysis of error scores, however.

Frequency-Amplitude Components of Tracking

It should be recognized that frequency and amplitude are, in the context of this presentation, aspects of the same phenomena, i.e., rate of angular change of the source of the tracking target stimulus. A report by Bowen and Chernikoff (4) indicates a predictive notion concerning the cumulative tracking error of subjects, by suggesting that an error estimate may be determined by the frequency-amplitude product, i.e., $E = kfa$. Tests of this relationship yielded good fits of the data. Although this relationship does not take into account the gain of the specific control utilized by subjects, there is significant indication of the "trade-off" between these two aspects of the tracking task to indicate reason why each should have pronounced differential effects and why there should exist an interaction between them.

Bowen and Chernikoff showed that equal frequency-amplitude products should yield equal tracking error. This point is not germane to the central topic of this report, but making comparisons of conditions having equal average velocities is essential to determining the applicability of the principle in the body of this research. So as not to confound the argument by possible interactions not revealed by previous analysis, matched groups for all subjects, presence, and complexity conditions were compared for the equal frequency-amplitude products included in the matrix of conditions presented.

Where $f_1 = 2^{-3.5}$ cps,
$f_2 = 2^{-2.5}$ cps,
$f_3 = 2^{-1.5}$ cps,
$a_1 = 2\text{cm}$,
$a_2 = 4\text{cm}$,

and E is the error associated with conditions dictated by the subscripts, the following hypotheses were tested by use of the t-test. The corresponding probabilities of the differences exceeding t are $H_1 : E_{f_1} a_2 = E_{f_2} a_1$, $0.3 < p < 0.4$ but $H_3 : E_{f_1 a_1} = E_{f_1 a_2}$, $p < 0.01$ and $H_2 : E_{f_2} a_2 = E_{f_3} a_1$, $0.5 < p < 0.6$ but $H_4 : E_{f_2 a_2} = E_{f_2 a_2}$, $0.05 < p < 0.10$. From the preceding, the frequency-amplitude equality concept can be considered to be valid. The differences between the equal fa products do not significantly differ, whereas cumulative error is compared for matched groups for like frequencies but different amplitudes. They do differ significantly or are near the critical level of accepted difference.

In addition to Bowen and Chernikoff's treatment of the frequency-amplitude product as a predictor of tracking error, Brown (7) has subsumed these components as those that may be translated into angular velocity, a primary stimulus to the human operator. Such a concept permits the translation of vehicle velocity and road curvature into just these terms. The stimuli employed ranged in maximum angular velocities from

0.74°/sec to 5.37°/sec. Under AASHO design standards for speeds 30-60 mph, on curves fitted to those speeds, angular velocities range from approximately 2.6°/sec for a 60-mph vehicle entering a curve whose radius is 1,330 ft with five percent superelevation, to approximately 6.7°/sec for a vehicle traveling at 30 mph on a 375-ft radius curve. Under these conditions the driver does not, in theory, need to make more than an initial correction, whereas subjects in this study were "negotiating a very crooked road indeed."

Although the differences between tracking error for equal frequency-amplitude products were not greater than could be expected by chance, there is a relative difference exhibited by the two amplitude conditions worthy of note. Inspection of the configuration confronting subjects shows at least two types of differences between the conditions: (a) pertaining to the angular velocity of control conditions, the average velocities for one condition tested differ almost 30 percent, but the maximum velocities did not differ; and (b) the maximum velocities differ by almost 10 percent, but the average differs only a few percent.

Implications for Signing

An earlier study conducted at the Institute of Transportation and Traffic Engineering at UCLA (2) revealed that as many as 15 percent of drivers made "poor selection of routes due to confusion of messages on the signs" or because "they had not seen the signs in time to make proper maneuvers." The driver is faced with a dilemma in negotiating high-speed roadways such as found in the interstate system. A wrong decision in selecting an appropriate ramp may take the driver many frustrating miles away from his desired destination. So it is especially critical to the driver who is unfamiliar with a specific course to "make the right decision." He may attempt to optimize the situation in such a manner that he has greater periods of time in which to reach a decision which is translated into reduced velocity, or he may direct his "attention" away from external visual sources such as signs; it is quite conceivable that his control performance, particularly lateral control, will be appreciably degraded.

This evidence from this study for the interaction effects of steering and search-and-recognition activities taken together with the report of ITTE suggests that the points where the driver must make decisions concerning alternative courses needs a substantial research effort. Congestion along the highway such as found at ramps does not explain the high accident records found in these areas. A report by Mullins and Keese (17) shows that more than 20 percent of accidents in freeway sections studied occurred at ramps where decision processes must be made on the basis of sighting entry into acceleration and deceleration lanes where impinging traffic and signing must be searched out, located or recognized and acted on. If recognition is to occur in comparable periods of time for decision points, then sampling of relative position on the roadway must be decreased and this is indicated not to be the case; or the angular velocity of the target (roadway) must be decreased. This may be accomplished by a decrease in speed, a factor that appears to be critically associated with accident rates (18). Where speed variability is high on high-speed roadways, there is also some evidence of a reduction in volume.

The issues that must be considered then are alternative methods of presenting information to the operator so that he is not overburdened, and ways of maintaining low variability of speed for such decision points along the highway.

A final point germane to signing practices and admittedly a general one is that where expectancy of a message unit is relatively great, an indication of the proper course should be shown on a sign, i.e., otherwise longer periods of time are presumably drawn to the sign-reading task at the cost of maintaining stable steering control.

SUMMARY AND CONCLUSIONS

Previous research on the independent aspects of tracking and search-and-recognition activities have not considered the interactions between these two tasks. Although studies of time-sharing between different sensory modalities and two-dimensional tracking have indicated generalized degradation of performances beyond certain input loads, consideration of tasks having different basic functional properties has been lacking.

The reported study did investigate the interaction processes between two such tasks and compares the alterations in performances where tracking and search-and-recognition performances are measured in a non-time-sharing role.

Analysis of individual performances revealed that the process remains essentially independent for each particular encounter with a time-shared situation, but tests of the specific hypotheses revealed a change in operator-pacing as the complexity of the activity increases. However, in all comparisons of the independent tasks and their time-shared counterparts, the presence of the second task yields a decrement in performance. The first hypothesis stated that tracking error would increase in proportion to the increase in each of the components of angular velocity (frequency and amplitude of a sinusoidal function). This hypothesis was borne out in part. Although the level of the tracking error was raised above the control tracking conditions in the presence of the search-and-recognition task, no differential increase occurred as a function of the angular velocity components. This was implied to suggest that the rate of sampling of the tracking signal was increased with its speed.

The effect of increasing the number of message units on the simulated sign-reading task had no effect on tracking performance, indicating an optimizing of time pacing where the greatest number of message units appeared in the search-and-recognition task.

The second hypothesis relating the effect of tracking frequency and amplitude on search-and-recognition speed and accuracy was not borne out. Recognition time actually decreased for the higher frequency tracking signals, but the effects of amplitude were less drastic. This further substantiated the theory that the rate of sampling is increased at higher angular velocities of the tracking signal. The false alarm rate was low for all conditions and did not differ between tracking conditions.

The final hypothesis tested in this research was that no differences would exist in tracking performance where the search-and-recognition task was with and without the inclusion of a previously learned message. This hypothesis was not borne out, as tracking was degraded where the key message was not included in the task. It is presumed that greater periods of fixation were required where the expected message did not appear, hence distracting from the visual response required for minimizing tracking error.

ACKNOWLEDGMENTS

The authors wish to thank members of the Bureau who acted as subjects during this study and especially gratitude is extended to members of the Electro-Mechanical Research Branch of the Office of Research and Development for their aid in developing a complicated facility for Human Factors laboratory research.

REFERENCES

1. Adams, J. A., "Human Tracking Performance." Psychol. Bull., 58:55-79 (1961).
2. "Driver Needs in Freeway Signing." Automotive Safety Foundation (Dec. 1958).
3. Bahrick, H. P., and Shelly C., "Time Sharing as an Index of Automatization." J. Exp. Psychol., 56, pp. 288-293 (1958).
4. Bowen, J. H., and Chernikoff, R., "The Effects of Magnification and Average Course Velocity on Compensatory Tracking." URL Rept. 5186 (Aug. 1958).
5. Briggs, G. E., and Howell, W. C., "On the Relative Importance of Time Sharing at Central and Peripheral Levels." U. S. Naval Training Device Center, Tech. Rept. NAVTRADEVCEN 508-2 (Oct. 1959).
6. Broadbent, D. E., "Perception and Communication." Pergamon Press (1958).
7. Brown, R. H., "Weber Ratio for Visual Discrimination of Velocity." Science, 131:1809-1810 (June 1960).
8. Desrosiers, R., Unpublished research conducted at the U. S. Bureau of Public Roads (1960).
9. Donders, F. C., "Die Schnelligkeit Psychischer Processe." Arch. Anat. Phys., 32:657, 681 (1868).

10. Ellson, D. G., and Gray, F., "Frequency,Responses of Human Operators Following a Sine Wave Input." USAF Air Material Command Memor. Rept. MCREXD-694-2N (1948).
11. Gain, P., and Fitts, P. M., "A Simplified Electronic Tracking Apparatus (SETA)." Aero Space Medical Lab., Wright Air Devel. Center WADC Tech. Rept., pp. 59-44 (Nov. 1959).
12. Gottsdanker, R. M., "Prediction-Motion With and Without Vision." Am. J. Psychol., 65:533-543 (1952).
13. Haynes, A. L., Mika, H. S., and Forbes, L. M., "Simulation Techniques in Automotive Human Factors Research." Paper Presented at Nat. Conf. on Driving Simulation, Santa Monica, Calif., Feb. 27 - March 1, 1961.
14. McRuer, D. T., and Krendel, E. S., "Dynamic Response of Human Operators." WADC Tech. Rept., pp. 56-524 (Oct. 1957).
15. Munn, N. L., "Psychology: The Fundamentals of Adjustment." Second Ed., Riverside Press (1951).
16. Merkel, J., "Die Zertlichen Verhatlnisse der Willensthatigkett." Rhilos. St. 2:73-127 (1885).
17. Mullins, B. F. K., and Keese, C. J., "Freeway Traffic Accident Analysis and Safety Study." HRB Bull. 291 (1961).
18. Solomon, D., "Accident Involvement on Main Rural Highways as Related to Speed and Other Characteristics of Drivers and Vehicles." Unpublished Rept., U. S. Bureau of Public Roads (April 1963).

Effects of Fatigue on Basic Processes Involved in Human Operator Performance

I. Simple Vigilance and Target Detection

NORMAN W. HEIMSTRA, TRUMAN M. MAST, and LARRY L. LARRABEE
Department of Psychology, University of South Dakota

•THE PURPOSE of the present investigation was to determine the effects of exposure to several fatigue-inducing situations on performance on two perceptual tasks—vigilance and target detection. Both of these tasks are of importance in most settings where a human operator is involved. Vigilance, or monitoring behavior, is required when an operator must detect aperiodic stimulus changes in his environment, whereas target detection involves the location and recognition of specific aspects of a stimulus situation confronting the operator.

A considerable amount of research has been conducted in these areas during the past few years with particular emphasis on vigilance behavior. In these investigations a number of task and environmental variables have been studied. For example, studies have been concerned with the effects of such variables as rate of signal presentation, intersignal interval, signal magnitude, background noise, etc., on vigilance performance. In addition to task and environmental variables, the so-called subject variables have also received some attention by researchers. This class of variables would include motivational, personality, perceptual, or experimental factors.

Fatigue could be considered as a variable that would most appropriately be classified as a subject variable. Although considerable research has been conducted dealing with effects of fatigue on operator performance, results are frequently contradictory. In general, most findings suggest that fatigue has little or no effect on most performance measures (4). However, inasmuch as vigilance is a rather basic process that is sensitive to a number of variables, it may also be sensitive to fatigue. This is also true of target detection. Consequently, the investigation reported here was conducted to determine if these measures were sensitive to fatigue.

METHODS

Subjects

Eighty male subjects were used; 40 in the vigilance phase of the investigation and 40 in the target detection phase. Subjects in each of the phases were assigned to one of four groups—three fatigue groups and a control group. This resulted in a total of 10 subjects in each of the four groups in the vigilance and target detection phase. The subjects were all college students and were paid to participate in the investigation.

Apparatus and Procedures

Subjects that had previously been exposed to one of three fatigue conditions, as well as non-fatigued control subjects, were tested in either a vigilance task or a target detection task. The three fatigue conditions are described, as follows:

1. Mental Fatigue Condition. Subjects in this condition were required to work multiplication problems mentally for 4 hr before being tested in the vigilance or target

---

Paper sponsored by Special Committee on Driving Simulation.

detection task. The subjects were presented with a booklet containing the problems and were instructed to perform the multiplication "in their head" and to place their answer beside the problem in the booklet. The problems consisted of four- or five-digit numbers that were to be multiplied by a one-digit number. The subjects were under the impression that the booklets would be corrected and were urged to do their best. In general, the technique was somewhat similar to that used in other investigations concerned with mental fatigue (1, 3).

2. Driving Condition A. Subjects were required to operate a driving device for a 4-hr period. The task consisted of maintaining a model automobile on a moving belt by means of a standard automobile steering wheel. When the car was not properly maintained on the belt an electrical contact was broken and a timer activated. This permitted a measure of "time off the road" for each subject. Although this measure was not used in the present investigation, the subjects were under the impression that this was a critical measure and were urged to do their best.

3. Driving Condition B. Subjects also operated the device for a 4-hr period. In addition, they were required to monitor a pair of red lights that periodically increased in intensity. These lights were located approximately 7 ft in front of the operator at the end of the belt in the driving device.

In many respects the driving device was similar to that used by Suhr (6) in his studies concerned with the effects of a periodic refreshment pause on simulated-automobile driving performance.

The 4-hr period that the subjects were exposed to the driving conditions was selected on the basis of subjective reports by subjects that had operated the device in preliminary research. Although longer periods of time could have been used, it was felt that the sensitivity of vigilance and target detection performance to fatigue could best be evaluated when fatigue was minimal.

Vigilance Apparatus.—The subject, immediately after completion of a fatigue condition, was seated in a sound-treated room facing a circular 18-in. ground glass screen. The screen was located approximately 36 in. in front of the subject. Whenever a subject detected a faint flash of light that appeared on the screen, he was to flick a toggle switch. The light, or signal, was a faint, near-threshold, circular spot, 6 mm in diameter, with ill-defined edges.

Data were recorded for each subject over five 20-min trials. During each of these 20-min trials a subject was presented with 20 signals. A subject's score for a trial was the number of signals missed. Intersignal intervals ranged from a minimum of 4 sec to a maximum of 250 sec.

When seated in the vigilance test room, a subject was initially presented with 10 signals in a 1-min period. This was to insure that the subject knew where the signal was located and was familiar with the response required. Following the last signal presented in the vigilance session, the 10 signals were again repeated.

Because the primary purpose of the investigation was to determine fatigue effects, the total number of signals missed was a more meaningful measure than pattern of errors or vigilance decrement. Thus a reward system was used which would tend to maintain a relatively high level of motivation and consequently reduce errors in the vigilance task. A subject was paid $0.05 for each signal detected but was fined $0.10 for each missed signal and false response. A reward system of this nature has been shown to maintain vigilance performance at a higher level (5). Also, the short mean intersignal interval used would tend to reduce errors.

Target Detection Apparatus.—The target detection task involved the detection of an odd, or different, letter from a background of similar letters. The odd letter was considered as the critical target to be detected, whereas the other letters which were identical with each other, served as background figures. Three background conditions were used with 8, 16, and 32 letters. In each case the background letters were similar in shape to the critical target. Letter combinations were B-R, E-F, C-O, and X-Y.

Each configuration was presented to the subject by means of a slide that was projected on a screen. A total of 192 slides were used and consisted of 96 slides which actually contained a critical target and 96 "dummy" slides that contained only background figures. An equal number of 8, 16, and 32 figure configurations were presented in randomized order. The entire series of 192 slides was repeated so that a subject was exposed to a total of 384 configurations during a test session. Each of the configurations was exposed for a 2-sec interval and there was a 2-sec interval between presentations.

The subject's task was to determine if the configuration contained a letter that was different from the background letter or if it contained no different letter or critical target. The subject recorded his decision in a booklet.

RESULTS AND DISCUSSION

Subjects with previous exposure to the fatigue conditions tended to miss more signals in the vigilance task than the control subjects. Differences in the number of signals missed by subjects in the various conditions were tested for significance by means of the Mann-Whitney U-test. It was found that subjects exposed to the mental fatigue condition prior to testing in the vigilance task missed significantly more signals than subjects in the control group ($p < 0.025$) and in driving condition A ($p < 0.05$). This analysis was made on the total errors of each group for the entire vigilance session. Figure 1 shows the mean number of errors (missed signals) for subjects in each of the conditions.

It is apparent from Figure 1 that subjects in the control group (Con) performed better on the vigilance task than subjects in the other groups. There was very little difference in the performance of subjects in driving condition A (DC-A) and driving condition B (DC-B). In the mental fatigue group (MF) the fatigue effects were most apparent.

Although the primary concern in the study was with total errors and not with pattern of errors, data were analyzed to determine if a vigilance decrement had occurred in any of the groups. It was found that no decrement had occurred in the performance of control subjects or subjects exposed to the two driving conditions. However, a typical vigilance decrement did take place in the performance of subjects exposed to the mental fatigue condition before testing. This was unexpected because of the reward system used and because of the short mean intersignal interval used.

The mean number of errors made by subjects in the various groups in target detection are shown in Figure 2.

Statistical comparison of the number of errors made by subjects in the different groups showed that the subjects in the control group had significantly fewer errors than subjects in the mental fatigue group (p 0.01) and driving condition B (p 0.05).

It was possible for a subject to make two types of errors in the target detection task. A subject could report that no target was present when actually one was present (Type A error) or he could report that a target was present when there was no target (Type B error). In Figure 2, the two types of errors have been combined. However, analysis of the frequency of the two types of errors revealed an interesting pattern. It was found that the control group made significantly fewer of the Type A errors than did the

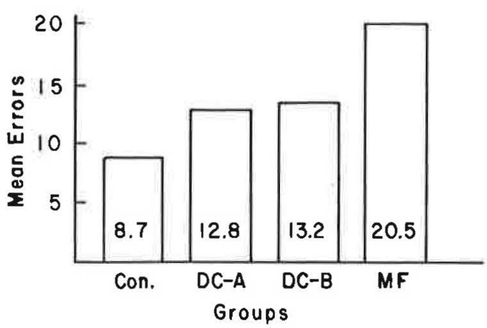


Figure 1. Mean number of errors in vigilance task for subjects in each group.

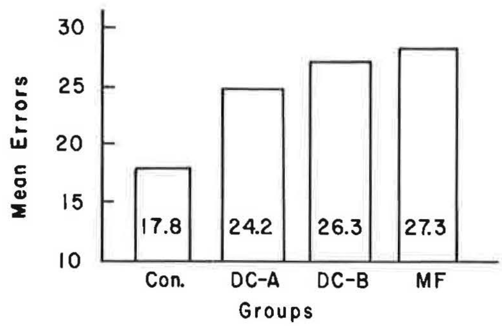


Figure 2. Mean number of errors in target detection by subjects in each group.

subjects in the mental fatigue group. However, differences between the other groups were not significant. In the case of Type B errors, however, subjects in the control group made significantly fewer than subjects in each of the other conditions.

Investigations dealing with fatigue ordinarily follow one of two types of procedures. In one case, subjects perform on a task and will continue, typically, for some period of time. During the session, performance measures are obtained on the primary task and possibly on subsidiary tasks. Ordinarily in fatigue studies of this type, subjects are assumed to be rested before beginning the session. A second type of fatigue study involves testing subjects, who have been fatigued before testing, on some form of performance measure. Examples of this type study are those that have been concerned with the effects of sleep deprivation on some performance measure. The study reported here also falls into this latter category. Obviously, both types of investigations can help in determining the effects of fatigue on driving. As pointed out by Crawford (2), the problem of fatigue in driving is twofold. It includes both the fatigue resulting from driving and the effects of fatigue, from whatever source, on driving. In most investigations concerned with fatigue and driving, however, studies of the first type, in which measures of performance were obtained as the subject drove an automobile or automobile simulator, have failed to reveal significant performance decrements.

In the present investigation, exposure to several conditions that were assumed to be fatigue inducing, resulted in modification of performance on both vigilance and target detection tasks. However, as is the case in most fatigue studies, it is difficult to state whether performance modification was due to changes in motivation of the fatigued subjects or due to more direct physiological changes.

CONCLUSIONS

Within the limitations of the conditions of this investigation, the following conclusions may be drawn:

1. Because vigilance and target detection performance are integral parts of operator performance, this performance will be most seriously affected by prolonged mental operations, which results in mental fatigue, and somewhat less affected by skill fatigue.
2. Vigilance performance and target detection performance both appear to be relatively sensitive to fatigue.

REFERENCES

1. Arai, T., "Mental Fatigue." Teachers' College Contribution to Education No. 54, Columbia Univ., 117 pp. (1912).
2. Crawford, A., "Fatigue and Driving." Ergonomics, 4:143-154 (1961).
3. Huxtaale, Z. L., White, M. H., and McCartor, M. A., "A Re-performance and Reinterpretation of the Arai Experiment in Mental Fatigue with Three Subjects." Psychol. Monogr., 59:5 (1946).
4. Silver, C. A., "Performance Criteria—Direct or Indirect." HRB Highway Research Record 55, pp. 54-63 (1964).
5. Sipowicz, R. R., Ware, R., and Baker, R. A., "The Effects of Reward and Knowledge of Results on the Performance of a Simple Vigilance Task." J. Exp. Psychol., 64, pp. 58-61 (1962).
6. Suhr, V. W., "Final Report on Study of Effect of a Periodic Refershment Pause on Simulated-Automobile Driving Performance Efficiency." HRB Bull. 212, 27-37 (1959).

Computer Simulation of The Automobile Driver

A Model of the Car Follower

MYRON L. BRAUNSTEIN, KENNETH R. LAUGHERY, and JOHN B. SIEGFRIED
Respectively, Flight Safety Foundation, State University of New York at Buffalo, and Minneapolis-Honeywell Regulator Co.

Driving simulation has generally involved an artificial representation of the environment and an interface for the presentation of the simulated environment to human subjects. For at least the better structured part-tastks, the development of a single system capable of simulating both the driver and the environment could eliminate this severe interface problem. This study is directed toward the development of a digital computer program of the information processing type which simulates the behavior of the individual driver in interstate highway car following.

Objective measurements and verbal reports were collected in a series of car-following runs on the New York State Thruway. A preliminary information processing model was prepared in flowchart form. Quantitative detail was added to the model using data extracted from existing literature and from a psychophysical experiment conducted on the Thruway. Programing and testing of the model is planned.

•DRIVING an automobile encompasses extremely complex patterns of behavior, involving sensory and perceptual processes, decision making, and psychomotor skills. Current knowledge of the effects of the variables influencing this behavior, and of their interactions, is limited. For many aspects of the driving task, the relevant variables have not even been identified.

These considerations suggest at least two approaches to the study of the driving task. An analytical approach is required to help identify the variables involved in automobile driving. Experimental studies are required to determine the effects of specific combinations of these variables.

In general terms, the ultimate objective of the analytical approach taken in the present research is a formal description of the driving task, in a form which allows a determination of the exact implications of this description, and of variations in this description. Thus, this research is directed toward the formulation of a model of the driving task. The model should assist in the selection of a manageable subset of variables for experimental study, as well as improve general understanding of the driving task. This paper describes a methodology used in modeling driver behavior as applied to the car-following part-task.

The driving task is of sufficient complexity to warrant use of the most powerful analytical techniques available. The approach taken in the current study is computer simulation. The reasons for selecting computer simulation, and more specifically the complex information processing (CIP) approach, are briefly discussed in the following paragraphs.

A computer program is both a flexible framework for the construction of a model and a convenient device for objectively determining the implications of the model. The program can readily combine conventional mathematical models, where these are

Paper sponsored by Special Committee on Driving Simulation.

postulated as approximations to observed behavior, with complex logical processes where these seem more appropriate. Theorized elementary processes may be represented as sub-programs and combined in a variety of larger programs to simulate various interactions of the sub-processes.

Perhaps the most important qualities of the computer model result from the formal structure required by the computer program. The program must be complete and logically consistent or it cannot be processed on the computer. Computers are designed with no tolerance for ambiguous instructions or faulty logic. Further, the entire program is available for direct inspection and is thus a complete and detailed record of every assumption in the model which it defines.

The specific techniques selected for the modeling of the driver's task, which will be referred to hereafter as the CIP approach, are those associated with Newell, Shaw and Simon (1). Some of the many applications of this approach are reviewed by Reitman (2).

Two distinguishing characteristics of the CIP approach are the use of verbal reports elicited from subjects performing the task to be modeled, and the programing of complex logical processes as contrasted to conventional mathematical functions. The collection of verbal reports is an old idea in experimental psychology. A tendency to rely on the content of such reports as reflections of internal states of the organism led to the abandonment of the technique by the behaviorist school, which has since begun to accept the reports themselves as data ("verbal behavior").

The CIP approach makes use of the content of verbal reports. It does not, however, use this content directly in answer to experimental questions. Instead, the reports supplement, and often unify, experimental data and analytical study in the formulation of a computer model. The model can then be tested for validity by comparing its predictions to experimental data. The value of the verbal reports is thus measured by their usefulness in the formulation of a valid model.

The use of logical processes in the computer model lends flexibility to the form of the model, as compared to approaches which are directed toward particular forms of mathematical functions. The form of the CIP model tends to stay closer to the hypothesized behavioral processes, which remain recognizable in the computer program. Revision of the model can more readily involve re-evaluation of hypotheses about the underlying processes, rather than parameter adjustment exclusively.

The CIP approach has generally been concerned with the detailed behavior of individual persons rather than overall measures of group performance. This emphasis, and an effort to treat all behavior as deterministic rather than include random elements in the model, has led to the inclusion of larger numbers of parameters than conventional models of equivalent behavior. This not only enables study of the sources of variability in the behavior modeled, but permits an evaluation of the effects of variables which do not even occur in conventional models.

CONTROLLED OBSERVATION OF CAR FOLLOWING

The part-task selected for study was car following, more specifically defined as maintaining a "safe and comfortable" distance behind a specified vehicle on a four-lane, divided, limited-access highway. A series of controlled observations were made of drivers performing this task on the highway. Both verbal reports and objective performance measures were collected.

Method

Apparatus.—A 1955 Buick sedan was used as the following vehicle. A bicycle-wheel generator-tachometer attached to the rear bumper provided a voltage signal to a meter inside the vehicle. The meter displayed the following-car velocity in miles per hour. A potentiometer connected under the hood to the accelerator linkage received its energizing voltage from the automobile battery. The output of this potentiometer was displayed on a voltmeter inside the vehicle, providing an ordinal measure of accelerator pedal displacement.

A motion picture camera in the following vehicle photographed the lead vehicle. A "target," consisting of two vertical poles and a crossbar, was mounted at the back of the lead vehicle. The vertical extent of the target on the motion picture film was used to determine the inter-vehicle separation. The same camera photographed the velocity and accelerator position meters, through a mirror mounted on the dash.

A tape recorder was used in the portion of the study involving verbal reports. A timing device in the following vehicle flashed a light in the field of view of the camera every 20 sec, and simultaneously sounded a buzzer, which was recorded on tape with the verbal reports. This permitted synchronization of the picture and sound records.

Procedure.—The observations were conducted on the New York State Thruway. Before entering the Thruway, the subject (S) drove the test (following) vehicle on local roads for approximately 6 mi. He then parked at the Depew entrance to the Thruway, where he was instructed to begin following the lead car, maintaining a "safe and comfortable" distance. If, while driving on the Thruway, the following distance was so great as to result in other vehicles cutting in between the lead and following cars, S was asked to select a distance which would tend to prevent these occurrences.

After a 10-mi practice period, test car velocity and accelerator pedal position, and inter-vehicle separation were recorded for 8 mi. S was then instructed to report any change he detected in lead car velocity, inter-vehicle distance or relative velocity, to explain any change in his accelerator control, his use of the brake, or the velocity of his car, and to give any other relevant information about his observations and thoughts. Verbal reports were elicited for the next 22 mi. For the following 8 mi, the verbal reports were recorded on tape and the objective data on film. The task analysis was based primarily on these last 8 mi.

Three laboratory technicians served as Ss. One maintained distances which led to frequent cutting in of other vehicles between the lead and test cars, and his data were not analyzed. Verbal reports were collected from two psychologists under conditions similar to the test runs and those of the more verbal one were used to supplement the data obtained with the two technicians.

Results

The verbal reports were summarized and categorized to provide an indication of the elements in the environment to which the subjects stated they were responding. Descriptions of their control actions and intents were also noted. Table 1 summarizes all verbal reports obtained from one or more of the 3 Ss whose reports were analyzed.

Tables showing the correspondence of the verbal protocols to the objective data, for 5-sec intervals, were prepared for 2 Ss. Table 2 presents a 30-sec example of these data for the S who gave the more detailed report.

A wide range of observations was encompassed by the verbal data. There was considerable duplication from subject to subject, indicating that the small sample used probably covered much of the range of verbal behavior obtainable in this situation. A

TABLE 1

SUMMARY OF VERBAL REPORTS

Element	Observation
Lead car velocity	Increasing, increased, decreasing, decreased, constant, estimate (quantitative), description (qualitative)
Gap	Increasing, increased, decreasing, decreased, constant, estimate (quantitative), description (qualitative)
Test car actions	Accelerating, increasing gas, decelerating, reducing gas, leveling off speed, coasting, speed estimate (quantitative), matching speed to lead car's, closing gap
Test driver actions	Checking rear view mirror, checking side view mirror, adjusting rear view mirror, surveying landscape
Traffic	Passing cars, relative speed description for passing cars, vehicle which may cut in, behavior of passing drivers, vehicle following test car, projected action of following vehicle, no following vehicles
Road scene	Good view of road, poor view of road, curve, grade, overpass, bird crossing road, roadside maintenance, service area, exit
Signs	Caution (deer crossings), speed limit, distances to cities, service area, exit

TABLE 2

SAMPLE DATA TABULATION

5-Sec Interval	Verbal Report	Distance (ft)	Test Car	
			Velocity (mph)	Accel-erator
1	I'm driving about 150 ft behind our lead car.	298	55.5	0
2	We're on fairly level ground at the moment;	278	56.5	0
3	I see ahead of us, however, that we will be dropping down a slight grade:	321	57	7
4	The next mile or two seems to be somewhat downgrade. The lead car is pulling ahead just a little bit and I'm acceler-ating to	315	61	2
5	close in the gap. I'm possibly doing	321	61	1
6	about 60 or 62 mph now; the gap is closed in and I'm leveling off.	315	60.5	0

difference in the fluency of the subjects was also noticed. One S gave several quantitative estimates of the gap, naming distances which were consistently about one-half of the actual distances.

The verbal reports contained numerous references to the lead car and to the gap. These references implied that the driver was responding differentially to these two elements, correcting for changes which had occurred in the size of the gap and adjusting his velocity to compensate for changes which were occurring in the velocity of the lead car. This distinction between responses to the lead car and responses to the gap is maintained in the model.

In general, the verbal reports, supplemented by objective performance indices, are sufficiently detailed and consistent to serve as the basis for a description of the driver's behavior in the car-following part-task. A computer model of this description was prepared in flow-chart form.

THE MODEL

The procedure in constructing a model is to work from the general, broad categories of behavior to the more specific behavior patterns. For example, in the present task, three general categories of behavior were identified: (a) selecting the direction of observation, (b) noticing an element in the environment, and (c) responding to the environment. Once these categories are identified, it is possible to describe the behavior in greater and greater detail until a level in the model is reached where elements such as the change in lead-car velocity or the content and location of a road sign are being described.

A detailed description of the model, together with the 22 flow charts which have been prepared, is presented elsewhere (3).

The main program, or executive routine (Fig. 1), sequentially executes three major subroutines. These subroutines determine the driver's direction of observation, specify the element in the environment which he notices, and produce a response to the noticed element. This three-subroutine cycle continues until a subroutine indicates that the run has ended.

The subroutine which determines the driver's viewing direction makes this determination on the basis of priorities assigned to elements in the various directions and acceptable time lapses between looking in various directions. The noticing of elements in the environment is dependent on priorities assigned to these elements by other subroutines.

A general response routine (Fig. 2) brings in an element response routine appropriate to the element in the environment which is noticed. The gap response routine is shown in Figure 3. These routines use the specific characteristics of the element, its momentary description, as the basis for selecting a response. Before a response involving velocity change is executed, a subroutine checks the acceptability of the intervehicle separation, according to criteria based on overall gap preferences and the momentary situation (Fig. 4). When a velocity change is called for, subroutines (Figs. 5 and 6) select the pattern of change according to the current velocities of the two cars, the current gap, the desired gap, and the time allotted for correction of the gap. The routines illustrated constitute a hierarchy of subroutines concerned with the driver's

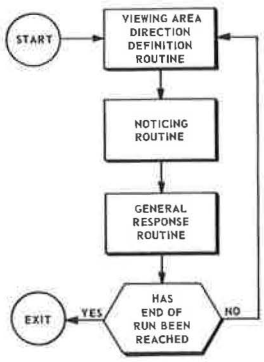


Figure 1. Executive routine.

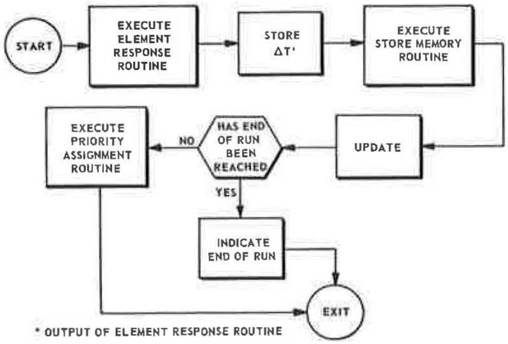


Figure 2. General response routine.

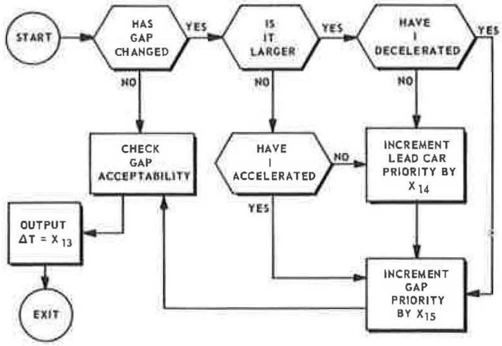


Figure 3. Gap response routine.

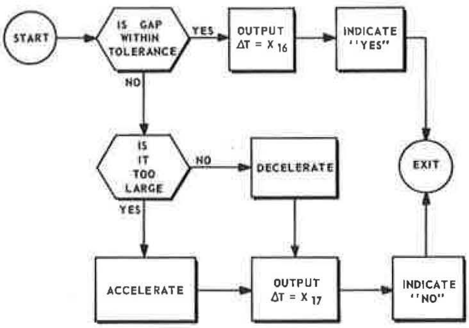


Figure 4. Check gap acceptability.

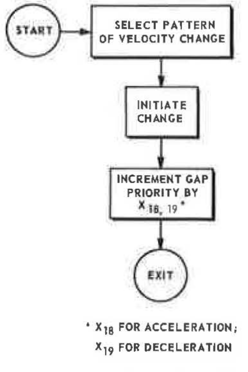


Figure 5. Accelerate (or decelerate).

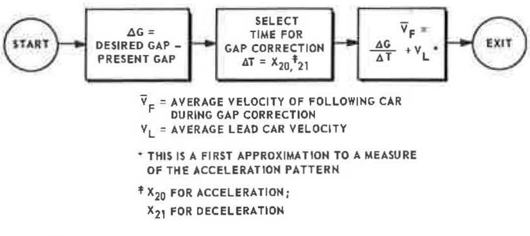


Figure 6. Select pattern of velocity change.

responses to elements in the environment. Other systems of subroutines are concerned with the selection of the viewing direction, noticing elements in the environment, and assignment of priorities to elements in the environment.

In addition to these routines, which attempt to simulate driver behavior and may be referred to as "psychological" routines (4), there are the "bookkeeping" or "non-psychological" routines which are concerned with updating the environment. These routines keep track of the external inputs to the driver and of the changes in these inputs which result from the simulated behavior of the driver.

The accuracy of the model in predicting the behavior of individual drivers under specific circumstances is dependent on the inclusion of a sufficient number of the relevant parameters and the determination of the correct values of these parameters. Most of the parameters are single numbers, concerned either with time intervals or with the priorities for noticing elements in the environment. Several are more complex and require special subroutines to compute their values. These are the decision functions and threshold functions.

Parameters

Perception and Response Times.—These are small time increments which would be difficult to measure but for which reasonable estimates could be made. An initial version of the model might use an average discriminatory reaction time, such as 0.3 sec, for the value of each of these time intervals. These values could be refined through further experimentation. The effects of varying these time increments over reasonable ranges might be explored in the exercising of the model.

The perception and response times which appear as parameters in the model are as follows: (a) noticing an element in the environment, (b) perception of no change in lead car velocity, (c) perception of change in lead car and selection of response, (d) observation of gap, (e) decision that gap is acceptable, and (f) initiation of gap correction.

Criterion Times.—A second type of time parameter, which might be expected to have a greater effect on the car-following part-task, is the criterion time. Criterion times are times allowable between various events and times allotted for or required for various processes, other than simple perceptions and responses. These times will also require experimental determination. The following effects of variations in these times on the output of the model should be of interest: (a) maximum time between looking ahead, (b) maximum between looking in the same direction, for directions other than ahead, (c) maximum between observations of gap, (d) minimum for reading a sign, (e) time allotted for correcting gap by acceleration, and (f) time allotted for correcting gap by deceleration.

Priority Increments.—In the model, the noticing of elements in the environment, and therefore the responses which are made, depends on the priorities assigned to these elements. These priorities are not directly observable and could at best be indirectly inferred from verbal reports as to which elements are eliciting responses. It appears necessary, for the present, to restrict priority increments to one or two levels. These levels would be somewhat arbitrarily assigned as the values of the following parameters: (a) lead car, if velocity has changed; (b) lead car, if gap change is attributed to it; (c) gap, if changed; (d) gap, if following car has initiated acceleration; and (e) gap, if following car has initiated deceleration. The element which is to have a priority increment is named first.

Other Numerical Parameters.—It is assumed that if the priority of an element in the present direction of observation exceeds a specified value, the driver will tend to continue looking in this direction. An arbitrary specification of this value would be required, such as minimum priority which prevents change in viewing direction.

A final parameter, which can be specified from available data, is maximum distance for reading a sign. This value would, of course, be dependent on the parameters of the sign (e.g., letter size, contrast) and driver characteristics.

Decision Functions.—Two major decisions required in the model are the selection of the desired gap and the selection of a pattern of velocity change, by the following driver. Suggestions as to the form of these functions may be found in the literature and in the data collected during the controlled observations. Most of the studies in the literature involve conditions clearly different from the part-task defined in the present study, and their results must be applied with caution. The controlled observation data, although taken under relevant conditions, were not collected for the purpose of accurately estimating parameters, and give only a rough estimation of their values.

Threshold Functions.—The major threshold functions in the model are the threshold for the detection of a change in the size of the gap and the threshold for the detection of a change in the velocity of the lead car. These thresholds are not readily disginguishable in the "real world" as the two changes tend to be correlated.

Laboratory studies of the velocity difference threshold have generally involved small objects, short distances and stationary observers (5). It appears necessary, therefore, to conduct research on such thresholds in the actual highway situation. In an experiment conducted on the New York State Thruway specifically to provide threshold data for use in the model (3), subjects were instructed to respond to velocity changes by a lead car initially traveling at 55 mph. Responses occurred in 4 to 6 sec to velocity increases of 6 to 9 mph and in 5 to 7 sec to velocity decreases of 3 to 6 mph. The design of this study is an example of the use of the model to specify areas requiring further research.

CONCLUSIONS

The situations in which verbal reports have been used in the development of computer models of behavior generally do not involve the occurrence of events requiring responses at intervals which are independent of S's behavior. In a preliminary study, Braunstein, White, and Sugarman (6) found that useful explanations of individual responses could be elicited from the driver while he performs a specified part-task. The controlled observation portion of the present study confirms the feasibility of collecting continuous verbal reports in a situation in which the need for response is determined by external events. The most fluent S talked during all but two 5-sec intervals during an 8-min run. There were approximately 50 distinguishable response classifications in the verbal repots of 3 Ss.

The major processes inferred from a study of the verbal reports, supplemented by the objective data, were described in flow-chart form. Sub-processes were also flow-charted, until a level of detail was reached for which flow charts no longer appeared useful. The preparation of a model of the car-following part-task not only proved feasible, but was also found to be a useful method of consolidating existing knowledge of the driver's task and of pointing to the direction in which this knowledge most requires extension.

In general, the applicability of the complex information processing approach to the development of a computer model of the task of the automobile driver was confirmed. The next logical step is the coding of the model and its exercise on a digital computer. Further experimental studies are required to provide better quantitative estimates of the parameters of the model. A validation study, comparing the predictions of the model to new samples of observed behavior, will be necessary before practical applications of the model are recommended. Finally, the extension of these techniques to other parts of the driving task should be explored.

ACKNOWLEDGMENT

This research was performed at the Cornell Aeronautical Laboratory under Contract CPR-11-8686 with the U. S. Bureau of Public Roads.

REFERENCES

1. Newell, A., Shaw, J. C., and Simon, H. A., "Elements of a Theory of Human Problem Solving." Psych. Rev., 65:161-166 (1958).
2. Reitman, W. R., "Heuristic Programs Computer Simulation and Higher Mental Processes." Behavioral Sci., 4:330-335 (1959).

3. Braunstein, M. L., Laughery, K. R., and Siegfried, J. B., "Computer Simulation of Driver Behavior During Car Following: A Methodological Study." Cornell Aeronautical Lab. Rept. YM-1797-H-1 (Oct. 1963).
4. Laughery, K. R., and Gregg, L. W., "Simulation of Human Problem Solving." Psychometrika, 27:265-282 (1962).
5. Brown, R. H., "Weber Ratio for Visual Discrimination of Velocity." Science, 131:1809-1810 (June 1960).
6. Braunstein, M. L., White, W. J., and Sugarman, R. C., "Techniques for Determining the Requirements of Part-Task Driving Simulators: A Methodological Study." Cornell Aeronautical Lab. Rept. VK-1719-E-1 (Oct. 1962).

Development of the KAKEN* Driving Simulator

MINORU KOBAYASHI and TSUNEAKI MATSUNAGA
Traffic Safety Laboratory, Scientific Police Research Institute, Tokyo, Japan

•THE NECESSITY of synthetic research in the area of traffic is rapidly growing, as indicated by the major topics at the Williamsburg Conference (1) and National Conference on Driving Simulation (2). As is well known, the automobile driving situation consists of three major factors: driver (man), car (machine) and environment; these interact in a complicated way in both time and space. In view of this, a system research approach is necessary to find out the relationships between stimulus and response, or between input and output. Therefore simulation techniques must be included as one of the promising methodologies in human factors research (3).

Undoubtedly, the car driving situation is a kind of man-machine system which is quite different from systems such as the task of flying by instruments or a simple vigilance task where environmental factors, including noise, play a minor role among the components. On the contrary, driving behavior is quite "unprogramed," and environmental factors in this case will always include large amounts of uncertainty and unexpectedness. Thus, in the driving situation, the use of the term "man-machine processing system" is proper (4).

The greater the complexity or number of uncontrollable factors in the system components, the greater will be the usefulness and importance of simulation techniques and modeling.

DESIGN PROJECT

The importance of simulation techniques has been fully discussed by Hulbert and Wojcik and Fox (5, 6).

In 1959, when the Traffic Safety Laboratory was established at the Scientific Police Research Institute, the original plan to construct a driving simulator as a research tool was proposed. In promoting the design project there was concern with performance, cost, and maintainability (7), and an acceptable balancing point was found with the following items (Figs. 1, 2, and 3):

1. A real car with a simulated environment technique is used for achieving the maximum degree of feeling for driving.
2. For the visual display system a motion picture technique was adopted; the projection area will cover approximately 50° horizontally.
3. For ease and economy of operation 16-mm films are used.
4. A feedback mechanism between car speed and film speed is to be constructed.
5. Handling torque and a self-returning function are provided.
6. A feedback mechanism between the angular displacement of the steering wheel and the light direction of the projector is actuated by a power servo-motor.

VEHICLE DYNAMICS

The KAKEN* Driving Simulator consists of three major parts: vehicle dynamics, visual display system, and recording system. A passenger type car (Toyopet Corona, 1500 cc) was selected for the simulator car for ease of remodeling and because of the limited space of the simulator room. The main remodeling changes were as follows:

Paper sponsored by Special Committee on Driving Simulation.

*Japanese abbreviation of the Scientific Police Research Institute.

Figure 1.

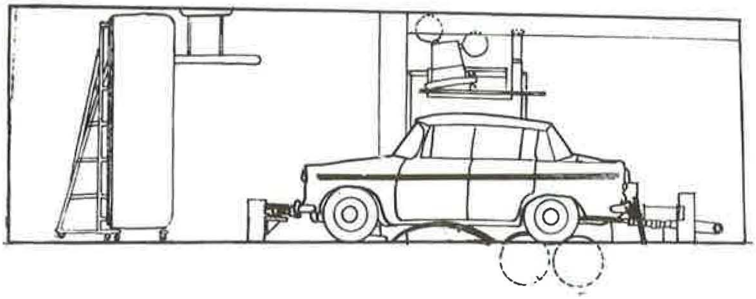


Figure 2. A cross-section of KAKEN driving simulator.

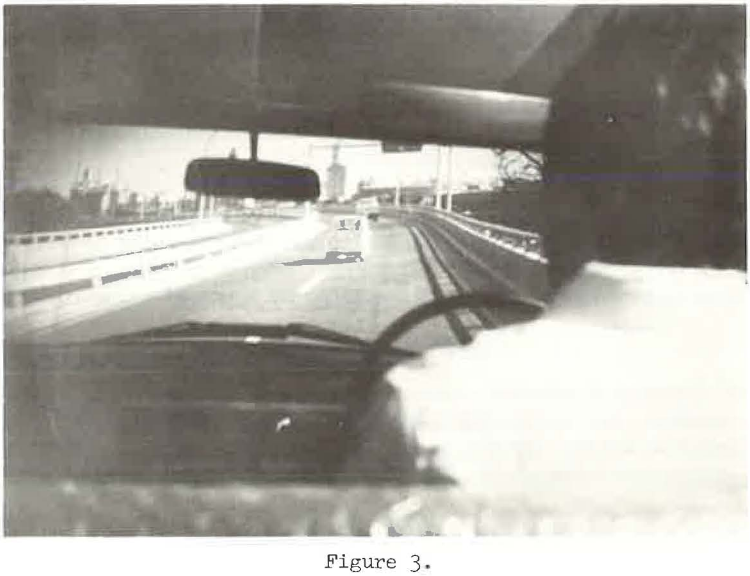


Figure 3.

1. To simulate the self-returning function of the steering wheel, the steering system is spring loaded.

2. To give drivers the actual feeling of "driving," and to avoid some transitional phenomena of the simulator car, special bumpers are attached at the front and rear ends, which absorb the reaction forces on starting and stopping.

3. A forced cooling system is used for the radiator and a continuous water flow is circulated in it.

4. To avoid risk of CO in the simulator room, a refueling pipe (50 mm in diameter) is connected to the exhaust pipe, and gas is guided to the outside of the building. Draincocks are also attached.

5. Intensities of the headlights, rear lights, and flashers should be reduced to an optimal brightness. Also, a low frequency oscillator will be used for the horn.

A twin free-roller system is used to transmit the torque from the rear wheel revolutions with maximum efficiency. In an attempt to achieve the highest fidelity for the system, the radii and widths were carefully calculated according to the characteristics of the simulator car (inside radius—0.24 m, width—0.80 m). Tensile strength is enough for driving at 100 km/h (62 mph).

VISUAL DISPLAY SYSTEM

The aim of introducing the motion picture technique here was to get high simulating fidelity and to reduce maintenance cost.

Actually, an analog simulator with a cathode ray tube like General Motors (8), or the miniature projection technique by the Institute of Transportation and Traffic Engineering (9) have interesting prospects, but they are still not practical. Also, a silhouette projection such as the SIM-L-CAR (10) seems inappropriate as a research tool, even though the visual feedback mechanism is superior to a film display.

For preliminary photographing of actual road conditions a motor-driver 16-mm movie camera (Bell and Howell 707) was provisionally set on the hood of the car, but test runs are being made to determine the best camera position. The projector is a Bell and Howell 7379, which produces few flicker phenomena at low-speed projection (Fig. 4).

Much effort was spent on the synchronization of car speed and film-feeding speed (number of frames per second) at the time of projection. As is well known, the feedback loop between human output and the environment plays an important role in a man-

Figure 4.

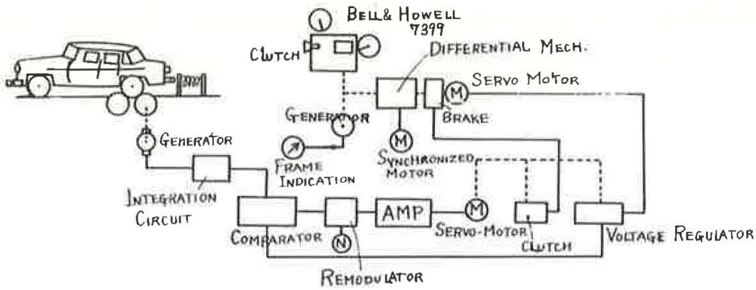


Figure 5. Block diagram of the synchronization of car speed and film feeding speed.

machine system. Especially, reproduction of the feeling of speed and acceleration is an indispensable factor in achieving high fidelity in driving simulation. Unfortunately, the film display is a programed system, but as far as possible an unprogramed situation is being maintained through this synchronization.

A following procedure is programed: When the simulator car starts, a voltage change is transmitted to the voltage regulator and actuates a servo-motor, balancing the output voltage. Then the servo-motor of the projector is actuated. When the car stops, the output of the voltage regulator tends to zero and a clutch starts to work. In this circumstance the projector works at a speed of only 8 frames/sec with power from a synchronized motor and gives the driver a feeling of very slow-speed driving. An almost linear relationship between car speed and film-feeding speed is achieved; e.g., 0 km/h = 8 frames/sec, 40 km/h = 32 frames/sec, and 60 km/h = 48 frames/sec (maximum speed) (Fig. 5).

To reduce cost and to satisfy the broad demands of the visual problem, a photographic speed of 24 frames/sec and a constant driving speed of 40 km/h was used. The superiority of this combination of projector and car speeds is discussed later.

To gain a wide visual field, an anamorphic lens (Prominar 16-A) is attached to a wide lens (1 in. focal length) and a projected area of 1.15 m (height) × 3.08 m (width) is obtained. At present, the feasibility of a superwide lens (0.5 in. focal length) is being tested. These lenses are usually used both in photographing and projecting and a curved vinyl screen is used. In the projection both sides of the screen are slightly out of focus but this situation follows the lack of focus of peripheral vision in normal driving.

RECORDING SYSTEM

The recording system is divided into three major parts: vehicle dynamics, physiological responses, and motion analysis. At present an integrated recording system between these parts has not been completed and they are recorded separately.

In respect to car dynamics, the car speed, braking force, accelerator pressure, steering wheel movement, and signal from the film are continuously recorded. As to physiological phenomena, galvanic skin response, electromyogram, and respiration rates are recorded. Driver's behavior (eye movements, hand movements, body sway) are under surveillance by means of closed-circuit TV and scored.

RESEARCH PROGRAMS

As mentioned earlier, systems research is a promising approach in road traffic analysis. The current research is concerned with the study of simulation fidelity through measuring human adaptability, speed, and distance judgements. A feasible study on skill decrements in the long range driving situation is programed.

FUTURE DEVELOPMENT PROJECTS

Linkage Between Wheel Movement and Visual Display System

Without the completion of the linkage between wheel movement and visual display system, a real "feeling" of car movements could not be obtained. In this regard, a

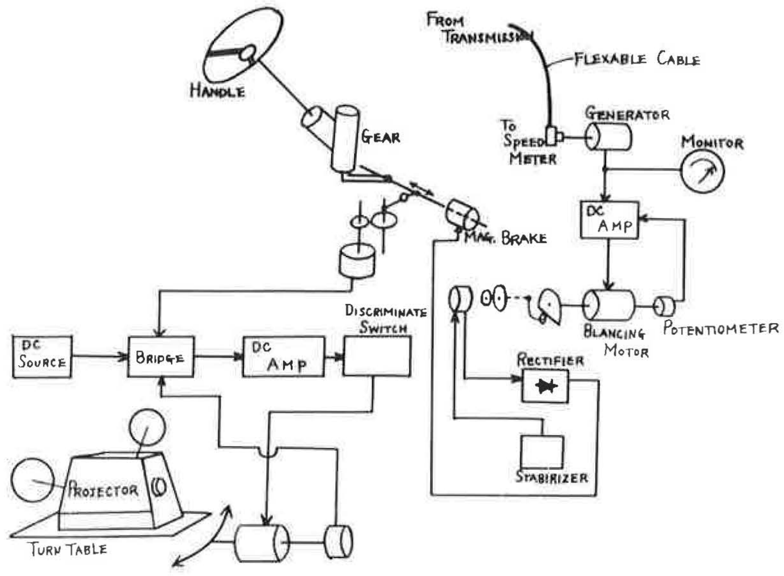


Figure 6. Block diagram of linkage between wheel movement and projector.

tentative linkage system is being designed which connects the angular displacements of steering wheel and the direction of the projector's beam through the electrical output amplified by servo-amplifiers. At the same time, variable handling torques are obtained for every handle position. The time lag of tracking performance for the turn-servo will be ½ to ¼ sec, and the range for the turntable movements will be ± 30° (Fig. 6).

Integration of Recording System

To analyze the interaction between the complex psychological and physiological phenomena and the traffic conditions, an integrated measuring and recording system is absolutely necessary. A remodeled 8-channel recorder, "POLYGRAPH" (Sanei Measuring Co.), will be furnished next year.

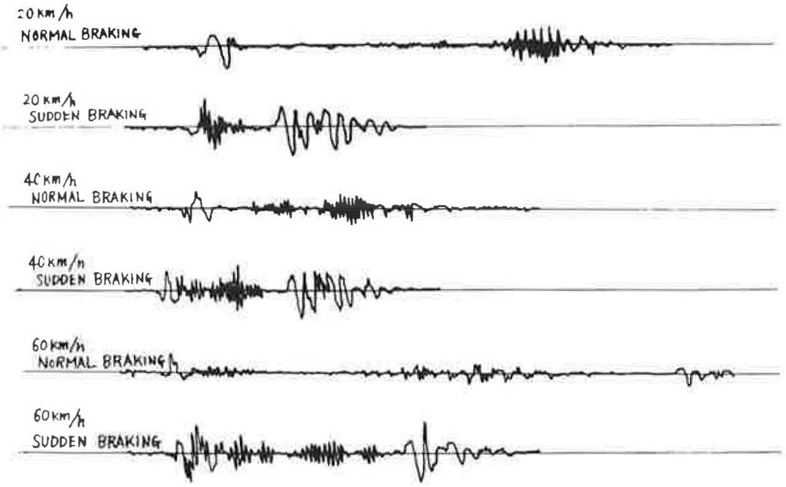


Figure 7. Fore and aft vibrations (rear sheet floor).

G-Forces Simulation

Because human beings are fairly sensitive to acceleration and deceleration forces, a carefully designed simulation system is desired. Discrepancy between the times of the visual stimulus and g-force will cause drivers to feel sick. As indicated in Figure 7, the present configuration of the driving simulator completely cancels the g-forces in the forward and lateral directions. These phenomena are due to the characteristics of the coil springs, and as a result, drivers are unable to feel g-forces. In this situation, inasmuch as this compensatory function in the forward direction was provided for, further simulation elements like UCLA's tilting chair probably will be added for the next project.

Although the presence of behavior dependent on illusion should be avoided in simulation techniques (7), a fairly large amount of body sway in the lateral direction by the "driver" on curved roads was found, even though the car was actually fixed. Clarifying the relationships between the amount of the driver's body sway, specification of road curvature, and car speed will be another interesting problem in human factor research. To some extent, the human induction mechanism toward the visual display will help the present research aim and will aid our judgment when we consider adding g-force simulation.

Rear Picture Projection

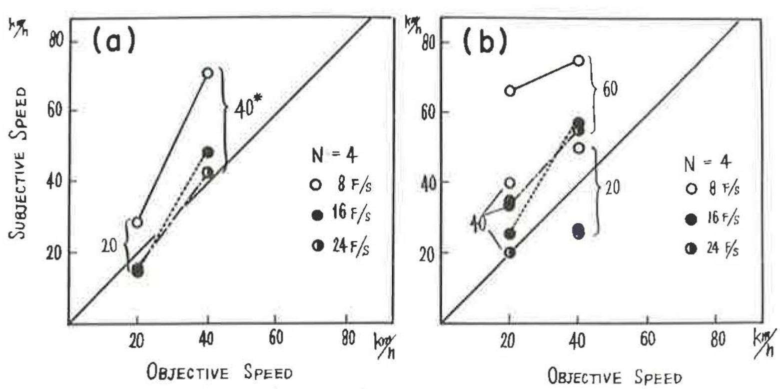
In car-passing or car-following behavior, a visual display system for the rear view would play an important role, but at the moment technical problems of film making and the continuity of experimentation do not permit pursuance of the problem. Presumably a careful study will be required to accomplish that objective.

AN EXPERIMENTAL STUDY ON SPEED JUDGEMENT

To find out the optimum film speed and car speed at the time of photographing, an experimental study was carried out. As a criterion car speed judgment by experienced drivers was studied.

First, a noiseless test road (a straight 800-m road with two lanes) was photographed at the Road Research Laboratory, Chiba Prefecture, with a 16-mm camera (anamorphic lens attached). Nine different experimental conditions (car speed—20 km/h, 40 km/h, 60 km/h × film-feeding speed—8 frames/sec, 16 frames/sec, 24 frames/sec) were tested. After arranging in a random order, the films were projected in the simulator room.

The experimenter drove the simulator car at a constant speed and asked the subject who sat next to the driver to judge the speed of the simulator car or film projected in front of him. The speedometer was viewed only by the experimenter (driver) in order



*Indicates car speed at photographing.

Figure 8. (a) Normal reproduction conditions, (b) different speed reproduction conditions.

to maintain a constant speed. The film speed and the speed of the simulator car were correlated as in the foregoing list.

Two experimental series (20 km/h and 40 km/h driving) were performed and the subjects who looked at the scene had to judge the speed in 10 km/h units (10 km/h, 50 km/h, etc.). Four subjects with long driving experience were used.

The film which was taken at 24 frames/sec showed the highest fidelity of reproduction for both the 20-km/h and 40-km/h speeds (Fig. 8a). In other words, the original car speed in the film taken at 24 frames/sec was estimated at the same speed in the simulated condition. The film taken at 16 frames/sec was fairly good, also. In a speed reduction condition (film taken at 40 km/h and projected at 20 km/h), a constant speed judgment was found in the film of 24 frames/sec. But both the 8-frames/sec film and the 16-frames/sec film diverged from the original speed (in Fig. 8b). In the case of increasing speed, the subjective judgements with the film at 24 frames/sec tended to be underestimated (Fig. 8b).

It was therefore concluded that the film photographed at a speed of 24 frames/sec had freedom and fidelity superior to the two other speeds. The car speed at the time of photographing was provisionally chosen to be 40 km/h for the present purpose.

ACKNOWLEDGMENTS

The authors express thanks to B. H. Fox for suggesting the presentation of this paper, and to S. H. Hulbert of UCLA for his valuable advice rendered during simulator construction. Also, many thanks to K. Suzuki, Director of the Traffic Division, Y. Ohkubo, F. Uruno and all of the staff in the Traffic Safety Laboratory and Traffic Control Laboratory, without whose constant cooperation this project would have been impossible.

REFERENCES

1. "Special Report on the Williamsburg Conference." Traffic Safety Research Review, 2:2, 1-11 (1958).
2. "New Horizons for Highway Safety Through Driving Simulation." Brochure, Nat. Conf. on Driving Simulation (1961).
3. Fitts, P. M., "Engineering Psychology." Ann. Rev. of Psych. 9, p. 285 (1958).
4. Tsubouchi, K., "Human Engineering." Vol. 24. (In Japanese.)
5. Hulbert, S., and Wojcik, C., "Driving Simulator Research." HRB Bull. 261 (1960).
6. Fox, B. H., "Engineering and Psychological Uses of a Driving Simulator." HRB Bull. 261 (1960).
7. Hutchinson, C. H., "Automobile Driving Simulator Feasibility Study." Cornell Aeronautical Lab., Rept. No. YM-1244-V-6 (1958).
8. Learner, D. B., "Development of the GMR Minimum Analog Driving Simulator." GMR 256, General Motors Corp. (1960).
9. Mathewson, J. H., "Progress in Driving Simulation for Research in Street and Highway Safety." TM No. L-13, Inst. for Transportation and Traffic Engineering (1958).
10. "SIM-L-CAR." Brochure of General Precision Systems Ltd., England (1960).

Some Criteria for Priorities of Research in Driving Simulation

Difficulties in Their Measurement And Application

BERNARD H. FOX and MARY W. FOX
Respectively, U. S. Public Health Service and George Washington University

•PLANNING FOR simulation research priority should be undertaken with some care because of the importance and potential value of products of such research, as well as its cost. Study of unimportant side issues or of poorly productive problems, wastes effort, time, money, and potentially, suffering and lives which might have been saved with a better-ordered research program.

Many attempts have been made in industrial and government laboratories to set up priority criteria. Where feasible, quantification has been applied. Mostly it has been possible to do this where payoff leading to realization of the objectives can be quantified; e.g., where payoff is money. In the military, the value of a weapons system is more difficult to measure. On the other hand, estimates of probability of success of the research can be made; this can be done more securely, the closer to engineering the task gets. In one sense, this case is similar to that of accident prevention, where a countermeasure is also contemplated. In the latter case, however, the countermeasure is not so clearly envisaged, nor can the fact of its identification, or of its achievement as a technique, be a measure of the success of the research. The probable success to be expected if it is applied must also be estimated.

There are various ways in which research originates. In other fields the origin of the research idea is similar to that encountered in basic research. In Weschler and Brown (20, p. 20, ff.), reference is made to the consumer; in the Navy to need formulation by the Chief of Naval Operations; to conferences; to technical levels within the organization; to occasionally high-level decision where a lack of knowledge or dearth of products may be known to exist; to previous work; and to results of an evaluation team. "The evaluation team itself, in fact, may be a research project" (20). Whether or not to do certain researches in the military is most often determined beforehand at command levels by a comparison of the effects of countermeasures, if they were to be developed. In accident prevention, however, the decision to do research is often based merely on the fact that something is unknown, without determining whether the knowledge derived from the research would have any value in reducing accidents.

But in whatever way a research is chosen, in the present field there is no sense in carrying it out without knowing its consequences. By its very name, accident-prevention research has an objective. If one does not measure against that objective, there is no point in doing the research. And, in fact, everyone who does research has said by implication, even if he does not formalize the evaluation, that he thinks there is enough value in the potential outcome to be worth the cost.

It is important, therefore, to have a measure of conformity to objectives in making decisions about research priority. This implies some estimate of the product of research, the ways the product can or will be used, and an estimate of the degree to which the outcomes of research will satisfy the objectives. In the present case, intuitive implication of payoff by a researcher before the fact is not dependable. There must be objective specification of estimated payoff before the fact.

Paper sponsored by Special Committee on Driving Simulation.

Weschler and Brown (20, p. 29) cite the criteria mentioned most frequently in a conference on evaluating research and development: financial—cost, profit, capital risk, utilization; feasibility and the need for the product; motivation of researchers; and schemes for evaluation of proposals. The California Research Corporation (20, p. 40) names these criteria: research aspects: probability of success (relating to research results only), research cost, time and other; economics: cost, payoff, investment, time, other; manufacturing aspects: operation; probability of successful commercial development (the last three are equivalent to implementation of a countermeasure in accident prevention); and last, overall appraisal. In all cases there were three levels for each measure. This is fairly realistic, in view of the uncertainties involved.

Most of the attempts at quantification have been made by commercial organizations, which have a much easier criterion to work with than government or other nonprofit organizations (10, pp. 6-8). One example (6, referring to Olin Industries, p. 22) is

$$\text{Project Value-to-Cost Ratio} = \frac{\text{Estimated Return}}{\text{Estimated Cost}} \times \text{Prob. of Success}$$

Several companies use analogous formulas.

Another example is Feeley's own statement of numerical rating of various criteria, resulting in an overall appraisal value (6). After applying an Index of Return formula

$$\text{IR} = \text{Total Estimated Return/Total Cost}$$

the IR is weighted by a factor involving a relationship (only shown graphically) of net profit to total future cost, which is then multiplied by an original rating assigned by a review board. Different projects are passed through this procedure and acquire relative numerical ratings, leading to priority selection. If payout period is very important, this may determine the weight assigned to IR instead of the factor relating profit to future cost.

In traffic safety, very few estimates of rank of priority have been made. For example, one of these described a numerical basis for rank (15). The data came from 944 returns on a questionnaire sent to safety educators, asking for opinions about certain research needs presented to them for rating. The basis of rank for a given project was weighted-sum-of-usefulness ratings. The projects were divided into survey and experimental study results. The highest ranking experimental (sic) study was "The characteristics of drivers with continuous records of safe driving." Unfortunately, the basis of the judgment of usefulness could not be determined from the results.

In general the problem faced in a simulation research program is the same as any research priority problem, but with the assumption that certain valuable tools are available in a given research facility. Where it may be possible to create new instruments, or to use other means of research (not all organizations can do on-the-road research), the problem becomes mixed. In that case, part of the question has to do with the relative priority of carrying out researches on the existing instrument, improving it, building another, or using means other than simulation. However, the problem of such decision-making can be subsumed under the general one of ordering research, inasmuch as the values associated with building new research instruments or applying specified techniques of research are included as criteria for research priority.

Some of the steps listed in the following were implied in the commercial formulas, but because the steps are often not specifically laid out there, one can only infer what the research director intended to assume or make estimates of, and what he intended to measure carefully.

The priority-determining steps to be applied to a list of researches are set forth in a statement of general approach with the express intent that the steps be not considered either as the only ones possible or as having only this sequence. Nor are they, as given, intended to be taken as final expressions of their best form, even if the particular steps mentioned are used in any given case. After the statement of general approach, particular formulations are given, followed by an illustrative example and a discussion of some of the considerations involved in the process of deciding.

IMPORTANT CONSIDERATIONS AND ASSUMPTIONS

Research List Sources

Lists of traffic safety researches are available from many sources. Some do not consider particular researches at all, but mention only general areas, or broad questions. Almost none names particular projects. Most describe mixed types of activity, some broad problems, and some particular projects.

A few sources of suggested research can be given, with some examples. Every issue of "Traffic Safety Research Review" (19) lists ongoing studies of many kinds and reflects the tenor of current activity. A list resulted from a 1956 conference on safety education (13). Two general areas were relevant: "general safety education" (59 research questions) and "traffic safety and driver education" (50 research questions). An example, randomly chosen from the latter, is "What predriver-education instruction should be given at the elementary school level and in junior high school?" Another list can be found in proceedings of a research correlation conference (18). From many suggestions, a special high-priority list was made up by the conference, consisting of 8 items. The eighth one happens to be of greater than ordinary interest for the present discussion: "As a research tool, development of an accurate simulator for driving." Other general lists can be found coming from various sources (9, 11, 14, 15).

Objectives

Objectives and some criteria may be the same. In this case it is assumed that the major objectives are the same as the major payoff criterion: the prevention of death and injury and the mitigation of injury and sequelae. They need not be the same. However, inasmuch as injury is usually associated with accidents, the researcher would therefore be interested in the prevention and mitigation of accidents as a means to these ends. Some organizations are principally interested in increase of basic knowledge. Some are concerned with efficient and inexpensive traffic flow, as well as accident prevention. Some are interested in research, whether in safety or in other fields, for training and teaching purposes. Some are mostly concerned with earning money through research.

As an example of a restriction based on the kind of objective, in this discussion only human factors in traffic safety are considered. Where there is mixed research subject matter, or doubtful content, a detailed analysis is necessary in the phase requiring assignment of values to research criteria. Meanwhile, researches obviously not concerned with human factors are not considered here (e.g., road material composition which reduces skidding).

Required Characteristics

A research activity must have certain qualities to be evaluated by anyone, whether they are directly stated or implied.

The possible outcomes of the research must be specifiable in advance as to type of information to be derived. As a necessary condition for this requirement, the degree to which an answer is not impossible (this is, in a sense, a kind of confidence statement about the reality of an observed change or difference) must be determined.

An additional condition demanded in the present approach is that all researches be carried to their logical conclusion. That is, it will be assumed that any research to be carried out will carry with it the expectation of a result aimed at payoff as previously outlined. The information to be gained (as a means of approaching payoff measure) is the result of the whole series of researches in a program or area. Thus, areas of research cannot be judged here until broken down into specific programs (or occasionally projects) with a stated sequence of researches and with known types of information emerging from the research series. When that is done, an estimate can be made of possible countermeasures and their probable success. By this means, diverse areas of research can be compared. But more, without the actual countermeasure, or at least a good guess at it, no payoff measure can possibly be estimated and hence no value attached to the research. It really makes no sense to say "We work in the area of alcohol and drugs because of their known importance in relation to accidents." The statement would better be expressed if something like this were added: "In this area, the following programs and projects: 1) ..., 2) ..., will give the following information: 1) ..., 2) ...,". With this information an estimate of effective payoff can be made. Without it, there is no idea whether the researches, whatever they are, would be any more than interesting intellectual exercises.

Moreover, certain other conditions must be fulfilled. The measure of desirability or worth of any research rests on the assumption that enough information is available about the research to carry out the evaluative process. This means that for any criterion named, a measure of that criterion should be available, however gross, however vague, with whatever uncertainty. Some important criteria are personnel, money, and time needed. To specify these, the researcher must have some idea of the design of the project. In addition, an estimate must be made as to whether a research can be done at all. Only after considering the method can this be determined. For example, in trying to tell how many traffic fatalities are actually suicides, it would be quickly established that the problem is very severe, and that a good solution—at least, one which produces confidence—is not easily evident.

This imposes a difficult task, inasmuch as at least a general approach to a problem solution must be available in order to make estimates of the criterion values.

Some researches to give examples of these considerations are, as follows:

1. Determine the validity of a particular driver-trainer for training purposes. In this case, relationship to injury and death is not clear, because in order to decide to consider the question as a research project one must first (or perhaps later) show that differences in validity lead to different payoffs. This has not been shown for any instrument. In fact, the only evidence available shows, in general, no difference between driver educated with and without trainers. This information has no bearing on the problem at hand. Thus, doing the required research alone leads to complete uncertainty as to the effect of the findings on payoff. The research cannot be put through the computational mill for priority because no value is given for the major criterion, payoff. This forces one of two actions: either abandon the research or put into the program a project or some means to establish the effect of validity on payoff. The outcome of this project would be guessed in advance (as with all unresearched projects), in order to assign a measure for purposes of priority valuation. Such guesses might, for example, come from a panel of sophisticated judges. (Would not determining accuracy of such judges be a valuable study?)

2. Examine characteristics of risk-taking behavior in the alcoholized driver. This question might be proper if given a little specificity. If there were an answer, it might reveal facts about increased or decreased tendency to make assumptions about the driving environment and the driver's car at different levels of alcohol among a large enough group to allow a feeling that the data had stability. Risk-taking would have to be specified by some measure, either for a task strongly resembling driving in its risk and hazard characteristic or one where the task is actually live driving.

No decision can be made on action resulting from the research outcome without making sure that increased or decreased risk-taking is associated with accidents. Therefore, a second research must form part of the program. It must be separated, of course, from the tendency to accidents due to other effects of alchol, and would probably be better examined by itself. Thus the problem is now: "Examine the effect of various levels of alcohol on a broad sample of drivers in respect to certain defined risk-taking measures. We assume associated research—not necessarily this project, but considered as a needed concomitant—on the effect of risk-taking on injury and death." If the research results lead directly to action and need no further information before recommendations are made (whatever they may be), one important requirement for listing as a research project is satisfied. The estimate of effect of risk-taking on injury and death can be made by many means, e.g., operations research, direct experiments leading to inferences, direct estimates from statistics. The estimate of judges about payoff, if used, is also required in advance of the study for priority estimation.

Thus, in using a payoff criterion of saving lives and reducing injuries, one is forced to set down the intermediate steps which must lead to the payoff. Otherwise one only deludes oneself if he is all for research and says vaguely, "We need to know the causes of accidents before we can suggest ways of improving the situation." Only in very particular, special cases can one assume an outcome whose payoff value is likely to be high no matter what the outcome of the research.

One might cite a negative finding as being valuable but even this must be useful only in terms of the remaining action alternatives always implied.

The matter of basic research is considered in the Discussion.

GENERAL STEPS IN PRIORITY SELECTION

The decision-making steps can be arranged in various sequences, but may follow somewhat along these lines:

1. Define the research tools, including simulator(s) and field facilities on which the research may be carried out. Implied is the restriction that researches will not involve equipment beyond what already exists except when the cost of new equipment can be included as part of a project. In defining the simulator(s) to be used as research tools, their input possibilities and the kinds of measures that can be made are described. These can then be reviewed when the limiting criteria are examined for elimination of researches to be done by a given technique.

2. List the limiting criteria—properties which make a project impossible regardless of other considerations: cost, time, negative payoff, equipment, personnel, mission, policy, etc. Define these in specific terms; e.g., eliminate all projects over X dollars or those falling outside the organization's mission. It is often convenient to carry out some elimination of projects while they are being listed for consideration, inasmuch as the limiting criteria are so evident and so binding, in many cases. This process is shown in the example.

3. Set forth criteria which can be used in evaluating priority of areas, programs, and projects not eliminated. Some of them may deal with the same characteristics as some limiting criteria, but in these remaining projects the value of the characteristic has not exceeded the limiting value. A good list will take much soul-searching, as well as much detachment. The detachment is needed in order not to miss certain criteria not immediately recognized as items determining research, such as public opinion or personal leaning. Soul-searching may come into the picture in assessing and setting down for all the world to see certain unpleasant criteria such as the director's or researcher's biases

4. Examine means of applying measures to the individual criteria. The word used here is "examine" rather than "determine" because it may not be practically possible, for some cases, that one can assign measure, or afterwards even make sense of it. By "measure" is meant a numerical assignment reflecting the degree to which the project has the particular characteristic in question. It is possible to find such measures with concrete meanings rather than to do so on abstract bases, as mentioned below.

The criteria here will be of three types: cost criteria (in money, personnel time, machine time, removal of teaching time, etc.); payoff (in a measure suitable to the objectives); and modifying criteria (which alter the value of either payoff or cost). Most considerations which are not obviously cost or payoff fall into the modifying category: e.g., risk of duplication of research by other organizations. In general, a payoff criterion P_i will be calculated by estimating maximum possible payoff P'_i with respect to a given variable and diminishing by attenuation with factors p_i relating to likelihood of finding and implementing countermeasures, their probable success, and the length of time before implementation. The cost criteria, C_i, are expressed in common units; likewise the payoff criteria. The modifying criteria may be either additive constants, k_i (payoff), and k^*_i (cost), expressed in the same unit as payoff or cost (depending on which is modified), with zero as the normal value; or as weights, w_i (payoff), and w^*_i

(cost), with one as the normal value to be applied to the gross cost or payoff scores as multipliers. These considerations will be expanded in later sections.

5. Determine means of combining criteria into payoff and cost measures: the payoff criteria P_i (in common units) and the cost criteria, C_i (in common units), are respectively summed to raw payoff and cost scores, P_G and C_G. Now the modifying criteria w_i are applied to obtain adjusted cost and payoff scores, C_A and P_A. Finally, to arrive at one overall priority rating for the project, there are two ways in which the cost score can contribute to this overall measure:

a. In the first method (method a), a weighting value, F, in terms of probable payoff per unit of cost within the program, can be applied to the adjusted cost score so that the weighted cost score will be in payoff units and can be directly substracted from the payoff. To the extent that this weighting factor, F, really estimates payoff per cost unit expanded, the weighted cost score, FC_A, is a proper measure of how much potential payoff is lost by the consumption of this amount of cost. Problems and techniques of obtaining such a weight are discussed later. The difference between adjusted payoff, P_A, and cost converted into payoff units, FC_A, will be the priority value, R_a, and projects can be ranked accordingly.

b. In the second method (method b), the cost potential for a stated period is specified corresponding to each of the cost measures stated in 4. All sets of projects, for which the sum of the cost measures does not exceed the cost potential of the organization (or any individual cost measure) are determined. In other words, all combinations of projects possible in terms of cost in the specified period are determined. If more than one simulator is available, and/or both simulation and other research are to be considered, then the various combinations must include all different ways of doing each project.

For each project in each set an adjusted payoff score is determined, just as in method a. The sum of the adjusted payoff scores for projects within a set results in a priority value, R_b, for the set as a whole.

Of course, combining of criteria can be done in a number of ways. The foregoing seemed reasonable on first analysis.

These schemes are described symbolically in Table 1.

6. Decide on techniques of applying the priority rating, R, to set up and maintain a program. Alternative 5b results in a program immediately by choosing for each simulator the set of projects which would be done in the order of priority value. However,

TABLE 1

Item	Payoff Measure	Cost Measure
Maximum possible payoff	$P'_1, P'_2, P'_3, \ldots P'_m$	
Attenuating factors	$p_1, p_2, p_3, \ldots p_k$	
Individual criteria (in comparable units, after applying attenuating factors)	$P_1, P_2, P_3, \ldots P_m$	$C_1, C_2, C_3, \ldots C_n$
Raw combined measures	$P_G = \sum_1^m P_i$	$C_G = \sum_1^n C_i$
Modifiers:		
Multiplicative weights (positive, with normal value 1)	$w_1, w_2, \ldots w_r$	$w^*_1, w^*_2, \ldots w^*_s$
Additive constants (positive or negative with normal value 0), and the same units as used for raw measure (e.g., deaths saved or dollars cost)	$k_1, k_2, \ldots k_t$	$k^*_1, k^*_2, \ldots k^*_u$
Adjusted measure	$P_A = (w_1 w_2 \ldots w_r P_G) + k_1 + \ldots + k_t$	$C_A = (w^*_1 w^*_2 \ldots w^*_s C_G) + k^*_1 + \ldots + k^*_u$

Weighting factor:
Priority Rating (method a), determined for each project and each research mode (field research, simulator No. 1, No. 2, etc.): $R_a = P_A - FC_A$
Priority Rating (method b), determined for each set of projects $R_b = P_A$, over projects in the set.

this may not always be possible. For example, the two projects with highest priority may be too expensive (in any of the cost measures) to do together at one time. A maximum priority combination within the cost limit can be found by method b. In some organizations the second project could be done part-time until the first is completed. However, if the organization is such that projects must be done in units of time and not on a continuing basis, the first and the second projects could not be done together and if it were decided to do the first and third, there is no certainty that that would be as good as the second and third or some other combination. The solution is provided by using method 5a in these types of organizations.

7. Define the operations which will provide numerical values of criterion measures. In general, the use of judges seems one reasonable approach at the moment, but the type and source of judges must be decided very carefully, because measures derived from their opinions are critical. Other approaches may involve operations analyses, which are minor or major research efforts themselves. The director alone may be the judge, or a panel from the organization or of outside experts in the field may be used. Different judges may be used for different measures. For example, it may be desirable to use outside judges to set raw payoff scores and judges from the organization to determine the cost and modifying criteria (e.g., policy factors). The type of average used to combine measures from several judges must be specified as well as details of the instructions to be given judges.

APPLICATION OF MEASURE AND PROCEDURE

1. List and specify the design of a set of possible researches as general areas, programs, and individual projects. Of course this implies, as mentioned above, a previous statement of objectives of research, which will set boundaries to the research list. Other boundaries will be imposed by consideration of nonobjectives and other limiting criteria. For example, an organization may be forbidden research which is another organization's prerogative. Let us assume that these objectives have been set forth and have formed a general basis for describing the list of researches. In effect, certain of the limiting criteria can be applied at this stage, saving the necessity of designing studies which will be eliminated later. But even then, often the mere research statement implies a method and an outcome, and permits a decision as to the applicability of the criterion measures. At any point in the process of design, the limiting criteria may be found to eliminate a project.

2. Eliminate all projects not meeting the limiting criteria.

3. For each project, obtain numerical estimates of criterion values on each simulator and other research mode.

4. Combine into priority ratings.

5. Determine program by methods decided on in part 6 of "General Steps."

6. Re-examine the whole picture periodically, based on developments, especially part way through projects.

EXAMPLE OF AN APPLICATION TO A PARTICULAR ORGANIZATION

Some details of applying these steps to needs in a hypothetical organization are discussed, together with actual estimates for two research projects.

1. Formulation of Measures and Procedure (definition of research tools—specifically, modes and organization, with application of a few limiting criteria in the process).

In the given case there do not exist the personnel or capabilities of running a field facility. In addition, only one simulator will be operable by the time the project starts. It will be capable of long-term runs; it will have acceptable visual resolution at distances greater than 25 ft; it cannot deal with intersections, interchanges, or turnoffs; it can interact with moving cars, both coming toward it and overtaking; short-term runs are likely to have low validity; etc. Only contract and in-house work are possible. Inhouse research is not feasible now. Within the contract type of project, only research to be done on the simulator is considered. Because this is contract work, available funds limit total personnel outlay, but do not limit the numbers at any given time.

2. Some Limiting Criteria Forcing Rejection of a Project.

(a) Cost exceeds "X" dollars.

(b) The research deals with road design characteristics.

(c) The research does not include work for a particular personnel specialty. (Otherwise stated, a particular person must be fired.)

(d) Total research time exceeds "X" months.

(e) The research deals with another agency's bailiwick.

(f) Simulator hours exceed "X" hours.

(g) Director's time exceeds "X" hours per week.

(h) Expectation of zero or negative payoff.

(i) Design involves curves in road, or intersections, eliminating use of existing simulator.

(j) Project does not give a result in terms of a factor such that death or injury could be reduced by eliminating, reducing, or changing the effect of the factor.

(k) Design requires short-run drives.

3. Some Possible Criteria.

(a) Payoff criteria.
   - (1) Reduction of deaths in the next 15 years as a result of the research.
   - (2) Reduction of injuries in the next 15 years as a result of the research.
   - (3) Not used here: earning potential for profit-making organizations.
   - (4) Not used here: training potential for graduate students.

(b) Cost criteria.
   - (1) Cost in dollars for the project.
   - (2) Not used here: cost in injuries or lives of doing research.

(c) Some possible criteria modifying payoff.
   - (1) Relative importance of preventing deaths and injuries in various population groups.
   - (2) Value of project for future research support.
   - (3) Biases of organization; e.g., policy considerations at various levels, consistency with previously announced research, attractiveness to researcher, and inertia of ongoing studies.
   - (4) Unpleasantness to experimental subjects.
   - (5) Training value.
   - (6) Public opinion.

4. Measures of Criteria.

(a) Payoff criteria. (P′. Estimate the annual number of deaths in the United States caused by—not just associated with—the factor(s) studied; e.g., alcohol under given conditions, ignorance of driving techniques.)

Then estimate the annual number of injuries in the United States due to the factor(s) studied and translate these into death units by multiplying by a proportion reflecting the importance of reduction of injuries relative to reduction of deaths. Some possible ways of specifying this proportion are as follows:

- (1) Use the ratio of number of injuries to deaths over the whole national accident picture. This figure would be changed every so often if the ratio changed appreciably in the population. The average figure is known from general statistics in the field (e.g., National Health Survey and National Vital Statistics data). If severity is included, each degree of severity could be applied to the ratio of the incidence of severity of injury to that of deaths. Care must be taken lest the objectivity with which a numerical value can be assigned to this ratio covers the fact that as a reflection of the relative importance of reducing deaths vs injuries, this value is very arbitrary.
- (2) Ignore minor injuries, and say that if deaths and severe injuries are highly correlated, deaths predict severe injuries and will measure payoff adequately. This assigns zero value to the proportion: importance of injuries compared to that of deaths. However, injuries would be reduced to the extent that they correlate with fatalities.

(3) Judges may assign an arbitrary equivalence ratio to death and injury. For purposes of simplicity, (2) is used in this paper; i.e., only deaths are used as a payoff criterion. Thus injuries are not estimated.

(4) Use accidents as a common measure instead of deaths or injuries, because it reflects both death and injury. With this measure, the differences in death and injury ratio, say between rural and urban experiences, are ignored. A ratio similar to that in (1) would result, but with property damage as a contaminant, while on the other hand, many unreported injuries would be included, as opposed to (2) and (3). Accidents are used here because of the complexity of evaluation required. Elsewhere they may be a good criterion measure.

The figure P', deaths per year attributable to the factor, is the maximum possible reduction of deaths due to the factor. To obtain an estimate of expected reduction in deaths due to the research in the next 15 years, estimate some attenuating factors. Some take the form of probabilities in the following group of a few examples:

p_1: probability that the project as designed will be sensitive enough to demonstrate the effect in question (assuming it exists).

p_2: assuming the effect is demonstrated, the relative contribution of this project to the total information (including that coming from this project) bearing on this factor, expressed as a value between zero and one, one indicating no other information on this factor. This factor should include consideration of the possibility that the project will be duplicated after it is begun.

p_3: probability of finding a countermeasure for the factor studied.

p_4: probability of implementing countermeasures, assuming one or more to exist.

p_5: expected effectiveness of countermeasures, assuming they are implemented.

These multiply to a raw payoff value for one payoff criterion. The other payoff criteria, if not combined, must be treated similarly to get raw payoff values for each.

The product of each P' and these attenuators must in turn be multiplied by a factor reflecting the 15 years minus delay in payoff to reach the final measure—that is, expected reduction of deaths in the next 15 years as a result of the research. The time, 15 years, is arbitrary as a base and could easily be 10 years or 20 years. Estimate y, the total time in years from the beginning of the research to the implementation of countermeasures. Thus, 15-y is the effective number of payoff years. Similar time factors apply to other P' criterion values. Thus the payoff measure for each payoff criterion, assuming more than one, is

$$P_G = [p_1 p_2 p_3 p_4 p_5 P'_1 (15-y_1)] + [p_1 p_2 p_3 p_4 p_5 P'_2 (15-y_2)] + \dots$$

in which the P' elements are maximum possible payoff associated with particular countermeasures and the y's are the estimated implementation times for these countermeasures.

(b) Cost criteria. (C = estimate of cost of the project in dollars.)

(c) Some criteria modifying payoff.

Each of these will be estimated as a multiplicative weight; i.e., a positive number with one the normal value, where values less than one are adverse and those greater than one favorable:

w_1: relative importance of the population group involved (e.g., children might be weighted greater than one).

w_2: value of project for future research support. (A study with value one neither enhances nor depresses potential for future support.)

w_3: relative importance of biases of the organization in relation to this project, including policy considerations of various levels, consistency with previously announced plans, attractiveness to the researcher, danger or unpleasantness to the subject, program training value.

w_4: relative importance of the expected public opinion about the project.

These weights have been limited to four factors here, but could obviously be expanded.

In assigning values to these weights the object is not to estimate the effect of public opinion, for instance, on the conduct of the project, but instead to judge the importance of the expected adverse or favorable opinion to management. Thus, if the project is expected to result in very poor public opinion, a weight of one may still be assumed if it is felt by management that public opinion is of no importance in this case. These weights can be interpreted in two ways: First, the effect of a weight of two (or one-half) will be to give twice (or half) the weight to deaths reduced by this study. Thus adjusted payoff is given by $P_A = w_1w_2w_3w_4P_G$. Second, a factor of two means that the factor is judged so favorably that this project is preferred to projects saving up to twice as many lives, other factors being equal. This last is a profoundly important aspect of the meaning of modifying criteria and is taken up in the Discussion.

5. Combining Criteria into Priority Rating.

For method a, applied here, an estimate of F, the probable payoff per dollar expended is needed. One approach would be to estimate payoff (based on the measure defined above) per dollar resulting from many researches already done, where costs are known. Then the geometric mean of these payoff-per-dollar ratios could be used as F. This value would be adjusted with passage of time as research expenditure per death and injury changed, or the value of money changed. The comparison is between several researches at any given time, hence momentary equivalence is required, and change of average cost of saving a life is permitted in later comparisons. Currently figures like a few dollars of research money per accident incurred in the population have been seen, but no figures for research money spent per life saved or injury prevented have been seen, either as a result of that research or of general research. In one agency, an estimate of $100, 000 used to build a certain type of highway is quoted as the expenditure necessary to save one life.

A second possibility avoids a difficulty inherent in the foregoing approach. Because research with high priority implies high payoff per dollar, the use of the average project to estimate payoff per dollar is likely to yield too low a value for this ratio. It might be better to use estimated payoff per dollar from a series of the most likely of the various researches being considered. This measures more directly the payoff value that would probably be gained from the dollars invested in this project if it were not considered among such a group. There must be an arbitrary decision as to how many of the highest priority projects to include in computing F. Also, before the projects can be ordered in priority, a "guessed" F value would have to be used—possibly resulting from the first approach. The modifying criteria should not enter into the determination of F. F should be the ratio of P_G to C, averaged geometrically over projects. Once the program is in progress, the value for F would be computed from the actual projects which have been recently carried out in the program. For illustrative purposes the geometric mean of the payoff per dollar from the two examples to be considered is used.

The measure is computed from

$$R = P_A - FC$$

6. Applying the Priority Rating to Set Up and Maintain a Program.

It is assumed that projects will be initiated in order of priority until the support potential is exhausted. New projects will be added as support will allow. Some projects can be done part-time to allow full use of money available.

7. Procedure for Obtaining Numerical Values for Criterion Measures.

Judges are used for all measures except cost, which comes directly from the design of the project. A group of "expert judges" will be chosen. They (or their staff) must know the traffic field. They must have skill in research design. They must be able and willing to get help in particular phases of the judging where they feel need of help. They must be aware of research procedures and costs in this field. They must spend the effort to do a conscientious job in spite of possibly not believing in the feasibility of this approach or the possibility of getting good answers to the questions posed. These expert judges will be asked to assign values to the P' criteria, the P_i factors, and y for each project. They will not be given knowledge of the method to be used in combining these into priority measures. They will not estimate w's except in special cases.

Instructions will be brief; the general purpose (priority ratings) will be explained with no further details. The values to be estimated will be defined as clearly as possible and illustrated. Consultation and use of as much helpful information as possible will be suggested. Extreme emphasis will be placed on judging values for every criterion to the best ability of the judge, in spite of the difficulties involved.

From the judged values for each judge, a P_G score—raw payoff score—will be computed and the average of these, over judges, will be used as the P_G value for the project. The payoff-modifying weights will all be set within the organization by the decision of a panel. For w_1 and w_4, the panel will consist of the program's policy makers; for w_2 and w_3, the researcher will also participate.

EXAMPLE WITH NUMERICAL VALUES AND APPLIED CONSIDERATIONS

1. Choice of Researches to Be Considered.

There are difficult problems associated with selection of research questions which are to be ordered in priority. It is therefore necessary to specify, in any technique devised for ordering, whether the researches are given and only have to be operated on; whether they must be conceived as problems out of the total field, and then operated on; whether they can be selected out of a known group and then operated on; or any combination of these. In the present case some approaches to the latter two conditions are indicated, and after a half-way selective process based on rational considerations, two research problems are stated very grossly and put to the technique.

An attempt is made to distill a narrow choice of projects for ordering from a large number of possible unspecified researches. The variables to be considered should be refined from gross areas to very particular variables. From the limiting criteria, short-term driving problems would not be highly valid. The simulator cannot handle intersection problems.

The most frequent conditions associated with injury and death on rural highways (this is where most long-term trips take place) include running off the road, alcohol, rear-end collisions, head-on collisions, and the pedestrian. In the case of alcoholic driving, the length of trip is likely to be short, excluding this kind of study. Running off the road, making up about 28 percent of the fatal and 14 percent of the injury accidents on rural highways (1), is a possible candidate. Under fatigue, a leading possible element in this kind of accident, probable long-term driving is found and fatigue is therefore a possible subject for study on this simulator. It will therefore be evaluated as a research variable. Other possible causes of running off the road are glare of headlights, "highway hypnosis" (a variety of fatigue) and sleepiness (also a variety of fatigue). All of these could be studied here, and are considered.

The case of rear-end collisions may be important. They make up about 16 percent of the rural accidents, and about 5.8 percent of the fatalities (1). This is a possible kind of study on the available simulator. It may involve any of the aspects of fatigue—sleepiness, tiredness, failure of vigilance, reduced perceptual skill, lowered performance skill, etc. In respect to the car, it may involve different configurations of rear lights, opposing glare, and the like. Pedestrians cannot be inserted into the simulator. In view of the limitations of money and time, and the acceptability as research areas of the several accident causes noted, these would be examined first.

One might consider stopping for rational reasons because such researches, sufficient in quantity, may have close to maximum probable priority and others will be unlikely to come close to the priority mentioned. On the other hand, what causes fatigue—reduced perceptual skill, sleepiness, etc.? It is for the very reason that they were selected to begin with—out of experience or intuition—that they would be considered high priority.

All of the variables mentioned have countermeasures; hence, that research on effects mentioned fulfills the requirements for consideration.

In this simulator, C3 can be eliminated (Table 2). A broad countermeasure resulting from simulator study is less easy to conceive for a permanent characteristic of the

TABLE 2

CONDITIONS AND THEIR COUNTERMEASURES ASSOCIATED WITH ACCIDENT BY TYPE

Item	Variable and/or Countermeasure		
A. Type of accident	Running-off-road	Rear-end collision	Head-on collision
B. Examples of conditions possibly associated with accidents not eliminated by limiting criteria, so-called immediate accident causes (2, 4, 7).	1. Inattention	1. Inattention	1. Inattention
	2. Sleepiness	2. Sleepiness	2. Sleepiness
	3. Tiredness	3. Tiredness	3. Tiredness
	4. Glare of headlights	4. Glare of headlights	4. Glare of headlights
	5. "Highway hypnosis"	5. "Highway hypnosis"	5. "Highways hypnosis"
	6. Drugs	6. Drugs	6. Drugs
	7. Skill decrement	7. Skill decrement	7. Skill decrement
	8. Perceptual decrement	8. Perceptual decrement	8. Perceptual decrement
	9. Unexpected curve in road	9. Unexpected curve in road	9. Unexpected curve in road
	10. Speed too high for conditions	10. Speed too high for conditions	10. Speed too high for conditions
	11. Skid	11. Poor perception on closing	11. Skid
	12. Others are not considered here		
C. Some other possible associated conditions, so-called intermediate and distant accident causes (2, 4, 7).	1. Age	1. Age	1. Age
	2. Sex	2. Sex	2. Sex
	3. Weather	3. Weather	3. Weather
	4. Experience	4. Experience	4. Experience
	5. Exposure	5. Exposure	5. Exposure
	6. Personality	6. Personality	6. Personality
	7. Driver training	7. Driver training	7. Driver training
	8. Personal characteristics	8. Personal characteristics	8. Personal characteristics
D. Countermeasures	1. Rest pauses	1. Rest pauses	1. Rest pauses
	2. Pep pills	2. Pep pills	2. Pep pills
	3. Attention-arousing or differently organized rear lighting	3. Attention-arousing or differently organized rear lighting	3. Attention-arousing or differently organized rear lighting
	4. Differential enforcement	4. Differential enforcement	4. Differential enforcement
	5. Alertness indicators	5. Alertness indicators	5. Alertness indicators
	6. Inside-the-car-radio reminders	6. Inside-the-car-radio reminders	6. Inside-the-car-radio reminders
	7. Sleep sidings	7. Sleep sidings	7. Sleep sidings
	8. Monotony breakers	8. Monotony breakers	8. Monotony breakers
	9. Unknown	9. Unknown	9. Unknown

driver, but seems to be more approachable for a behavior of the driver. Such a position is obviously subject to discussion, but for this case, C1 and C2 are dropped. Hypothetically C4 would be used, but these accidents are not, on the face of it, restricted to the young or old. They may contribute more than their share, but to determine why would involve too much of a research project for the present facilities. C5 is irrelevant to a simulator test. C6 could be tested, but it is extremely unlikely that this can be discriminated in such a small research project. Until one can get a sample of driver training groups who have not been selected by their volunteer status, this variable would better be avoided. C8 is possible, and could be considered. A more elaborate analysis could be made of C8.

Meanwhile, fatigue can be produced, possibly covering B 1-10, except 6 and 9, for running-off-road. The same holds for rear-end collision. If, in addition, one can get at B11, the range of this type of variable seems to be covered, even though many other intermediate, associated, or other dimensional variables are ignored. The choices (somewhat arbitrarily, it is true) are therefore personal characteristics, countermeasure evaluation, drugs, any of the separate elements involved in the fatigue process (such as B1, 2, 3, 5, 7, 8, and 10), B4 for all accident types, and B11 for rear-end collisions.

The countermeasure considered for falling asleep may not be the best one, but until one can find a better one, an existing one is best (D). Pep pills are already used, but so much abused that this cannot be used as a possible public countermeasure. Enforcement is irrelevant here in a simulation research. Alertness indicators thus far do not work well (but should not necessarily be discarded for another research). Radio reminders inside the car may work. Others are developing this work, however. Sleep sidings exist, but are not used enough. Monotony breakers self-imposed are not used much, but if someone invents a good external one it will be very useful. One cannot test it here because it has not yet been developed. There remain rest pauses, changed rear-lighting, and examination of as many of the items under B as can be studied incidentally during the test runs. The alternative is to study the items under B separately, the idea being that this information may permit better choice of countermeasure later, or even intensive research or development of a countermeasure not now particularly successful.

There are thus some countermeasures now to be examined. However, because it is known that many run-off-road accidents take place at curves and intersections, it may be better to emphasize rear-end collisions, which are less dependent on road configurations that this simulator cannot produce.

After going through much travail of the foregoing type, it may be possible to eliminate many more research possibilities and get down to a few. Some of the aforementioned reasons were given for examplary purpose, and not necessarily because of a tightly logical elimination process—see, for example, the consideration mentioned in the preceding paragraph. Assuming all the researches have finally been eliminated, the following two remain:

1. Misperception (not sleep) leading to rear-end collisions on turnpikes, with rear-end lighting countermeasures.
2. Drowsiness and sleepiness on turnpikes, with countermeasure a drowsiness-detecting and alerting device.

Project 1.

1. The research would consist of presenting several rear-end configurations in vehicle V_2, which is being approached at various speeds by vehicle V_1, in conjunction with different speeds of V_2, V_3, and their combinations. Countermeasures (e.g., rear-end lighting changes) would be pinned down, validations and cross-validations carried out, and programs initiated.

2. Limiting Criteria.

Most of these have already been mentioned in the foregoing discussion on choice.

3. Criterion Measures.

For Study I, several sources could have been consulted for probable number of lives lost due to the factor: the Pennsylvania Turnpike study (3), the Northwestern study (2), the estimates of some recent studies (e.g., 1, 12), and the like. A value of 3,000 lives/year was used as an arbitrary (probably not correct) number for P'. There is only one payoff criterion, P'. Thus p_1 will be the same as P_G. The attenuating factors mentioned here are p_1 through p_5. For p_1 it is not difficult to detect with confidence the effect of a moderate to strong countermeasure. The value of p_1 is therefore given as 0.90. This research, although important, will probably lean heavily on other parallel work. Hence $p_2 = 0.25$. Countermeasures are known. But even if not known, it is suspected that a highly effective one is feasible. It has been conceived, and no known objections prevent its application. Hence $p_3 = 1.00$. If the countermeasure is found effective, there is a good likelihood that it can be adopted, but not certainly, hence $p_4 = 0.80$. If implemented, the countermeasure might not be wholly effective because other elements enter the picture of rear-end collision based on misperception. The problem cannot be solved completely by this means, hence $p_5 = 0.50$. It is not expected that the countermeasure can begin to be implemented for 7 or 8 years. It should take 5 years or so to saturation. Hence, effectiveness within the 15 years is given as beginning at the midpoint of beginning and end of saturation, 10 years, with an effective duration, 10 years, with an effective duration of 5 years. Raw payoff can be computed now:

$$P_G = p_1 p_2 p_3 p_4 p_5 (P')(15\text{-}y) = (0.90)(0.25)(1)(0.80)(0.50)(3,000)(5) = 1,350$$

4. Modifying Criteria.

It is felt that for different laboratories or organizations, a large variety of modifying criteria exist, and only a few can be mentioned here, let alone dealt with. For any single group of researchers, however, the number is probably not particularly prohibitive. Criterion w_1, importance of study, is not biasing in one direction or another in this case. Hence, $w_1 = 1.00$. Criterion w_3, organization bias, is negative, because in this case there is some question about future problems with vehicle regulations, making countermeasures more sticky to deal with; therefore, $w_3 = 0.60$. Criterion w_4, public opinion, is not likely to favor or reject this more than other researches, hence $w_4 = 1.00$.

Now, adjusted payoff for Project 1 is therefore:

$$P_A = w_1 w_2 w_3 w_4 P_G = (1)(1)(0.6)(1)(1,350) = 810$$

Project 2.

1. The proposal would consist of first testing a theory about a means of detecting oncoming drowsiness in advance of its happening, and then, if it is correct, developing and applying a countermeasure. One theory for an approach to this already exists, with supporting data, and estimates can be made of its probable correctness and of the liklihood of developing and applying countermeasures. Tests of long-term driving would be conducted, sleepiness induced, the technique tested, validating and cross-validating studies run, and programs initiated. The method of detecting drowsiness in advance is to use a multiple-electrode device (16). From this, a cheaper detector would be sought, or one which indicated corollary information with the same end product. It would be made available and put up for sale. Among those who drive much, it might find a ready market. Among those who do not, it could be rented.

2. Limiting Criteria.
These have been dealt with in selecting the project.

3. Criterion Measures.
As in Project 1, an arbitrary value was selected for P' (again, only one payoff criterion). In the actual case a best estimate of P' would be made.

4. Modifying Criteria.
As in Project 1, the attenuators p_1 to p_5 are evaluated. First, $p_1 = 0.75$, because there is not complete confidence that the variable in question, discrimination of drowsiness, will be reliably detected if present. The present research is almost the only work of this nature being done, hence $p_2 = 0.80$. It is not likely that a countermeasure will be found; therefore, $p_3 = 0.30$. If it is found, it is not likely to find widespread use: $p_4 = 0.05$. If it is used, it must work well for it to be commercially useful, so $p_5 = 0.98$. The timing arguments lead to y of 7 years and $15\text{-}y = 8$. P' is estimated at 4,000 lives/year.

Now raw payoff for Project 1,

$$P_G = p_1 p_2 p_3 p_4 p_5 P' = (0.75)(0.80)(0.30)(0.05)(0.98)(4,000)(8) = 180$$

The modifiers w_i, which, as opposed to the p's, can exceed a value of one, are considered. Again, no bias attaches to drowsy drivers: $w_1 = 1.00$. Also, no positive or negative aspects of the work will affect future research support: $w_2 = 1.00$. The organization favors research on degrading processes in the driver, and supports attempts at overcoming them, hence $w_3 = 1.2$. Assume that the public will be slightly more favorably inclined to this than other research, and $w_4 = 1.05$, because the organization attaches importance to public opinion.

Now adjusted payoff,

$$P_A = w_1 w_2 w_3 w_4 P_G = (1)(1)(1.2)(1.05)(180) = 227$$

5. Measures of Criteria.
The research costs for Project 1 are estimated at $300,000 (probably understated), and those for Project 2 at $400,000 (similarly understated, perhaps more so). No estimate is made of countermeasure cost because both will involve commercial ventures, and the figures for probability of funding and implementing countermeasures p_3 and p_4 have, in essence, taken into account the costs of production. Where such costs fall to the researcher, or to a public agency in which there is direct cost, not investment, they may be estimated. This portion of the problem requires much consideration (see Discussion).

6. Combining Criteria.

Ordinarily, F, the payoff/cost ratio, would come from a number of high-priority researches. But because there are only two here, they were used. F is obviously common to all researches being ranked. Here F is derived from P_G, not P_A, because the payoff is independent of w_i; e.g., organization bias (although possibly not of judges' fallibility), which only affects the researcher's payoff value, not the true payoff, P_G. Thus

$$F = \left[(P_{G_1})(P_{G_2})/(300,000)(400,000)\right]^{1/2} = 0.0014$$

Remembering that the priority of any research is expressed as

$$R = P_A - FC$$

separate values are derived for each research in question:

$$R_1 = 810 - (0.0014)(300,000) = 390$$
$$R_2 = 227 - (0.0014)(400,000) = -335$$

Hence, Project 1 is preferred to Project 2.

DISCUSSION

The present paper is felt to have particular use—whatever other values it may have or lack—in emphasizing the existence and importance (unrealized in many cases) of certain criteria in making decisions about what research to do. However poorly one may regard the validity or even the feasibility of the present approach, if one goes through the motions of estimating values for the various criteria, certain assumptions made about relative value come startlingly into view.

1. In estimating P', extent of the problem, one must examine just how much of a dent would be made if the problem were solved and a countermeasure implemented. For certain problems P' is quite unbelievably small. Only by specifically asking and answering, or making a best guess if the data are not available (as is usually the case), does one get to a feeling for one P' compared to another.

2. In determining p_1, values must be included for various elements of research which ordinarily are considered under evaluation of quality of research. Among other things these include researcher quality, motivation (20, p. 37ff.), excellence of research design, proper research team makeup, adequate statistical planning, and research know-how. At a secondary level these may themselves be evaluated by a process involving p's and w's. They have not been dealt with in this way for reasons of space, but the procedures can easily be seen in this context.

If one were asked specifically, it would often not be obvious that research quality of much of the literature is poor. It takes one kind of sophistication to know that one is knowledgeable. It takes another, equally important, to know that one does not know. One major value of carrying out these procedures may be that it might lead a researcher to question his own research competence (or that of his team) and to try to get an external evaluation of it. This particular field—accident prevention—has many examples of utterly useless projects because of this factor (echoed in 5, 9). The sad part is that because of the lack of insight, some people will, even if exposed to the necessity, be reluctant to put their competence to the test; and, if not so exposed, will not know that their research products are worthy of very little confidence.

3. For p_2 one may ask, as well as its major intent, about duplication of effort, and the likelihood that information developed by others will form the central basis of the outcomes of research. A fundamental element is how much the information being developed is needed for solution of the problem. If it cannot be solved at all without the

research, $p_2 = 1$. If an approximate solution cannot be obtained, p_2 has a lesser value. A difficulty arises when several researches are necessary, but none is sufficient. Does this necessary piece of work still retain its value of 1?

4. The values for p_3, p_4, and p_5 involve making an estimate of the existence, effectiveness, and probable implementation of a countermeasure. A process of evaluation must go into a value for countermeasures as well as for research. Note that in a fully developed scheme for research, the cost of countermeasures should be included, not only as to payoff, but also as to time and money cost. This consideration is not impossible to develop, but is quite complicated.

5. An important factor here, hardly to be overemphasized, is that when one is forced to estimate the value of the w's, one is doing so in terms of the payoff, lives. In this case, the question resolves itself ultimately to: by what factor is one multiplying the expected number of lives saved (after the p's are applied) in order to take into account the w's? If P_{G_1} is 500, and P_{G_2} is 550, and researcher preference leads to values of $w_{3_1} = 1.4$ and $w_{3_2} = 1.0$, then the researcher equates his positive bias to 150 lives. He is saying, "If I had two researches with the given values, I would do one where $w_{3_1} = 1.4$, since the effective criterion value is 700 in the one case and 550 in the other." Because the preference factor is applied to lives, in the long run, if these estimates of P_{G_1} were to be correct, the meaning of researcher preference or of the other w's would seem on the face of it to be impossible to gage—lives vs whims or administration policy. In the actual practice of research choice, however, this consideration is almost never seen as an equation of lives with policy or with preference, etc., by the researcher Calica et al. (4) specifically note this point. It is more often clearly seen as such by the countermeasure practitioner, such as the road builder, the traffic engineer, and the police administrator (4), who are forced into the recognition of their decision processes by realistic need. The same judgment should govern the decisions of research, given the stated objectives.

6. And so with the other criterion elements of w_3. Among them must be included inertia of the organizational structure. It is messy and personally distasteful and troublesome to replace people; e.g., two engineers testing seat belts might be replaced by two socioligists, who could possibly find out the why's of poor seat belt use and might come up with a countermeasure, or vice-versa, where a psychologist is trying to develop a driving licenses selector with a paper and pencil test. In another criterion case, administration may be against involvement in a particular research area. This refers to research areas where no absolute stricture exists, because where one does exist, it is regarded as a limiting criterion. The bias of administration is often treated as a limiting criterion if the researcher does not even consider bucking such a position. One kind of administration bias results from the probability that research support will be radically reduced if a glamorous effort is given the normally lesser priority in terms of lives that may be saved. Here one must look to administration for decisions on the weight to be attached, inasmuch as a balancing of ultimate outcomes is involved. Often the priority of the glamorous work rises not because of bias, but because realistically, ultimate payoff is then greater for that effort than any other.

Many other criteria can be thought of. When conceived, they should enter as basic payoff or cost criteria, or as modifiers with appropriate weight.

Another way of emphasizing the full meaning and impact of modifiers such as the w's is to ask, what weights would the w's have been given if the number of deaths due to traffic accidents were 400,000 per year instead of about 40,000 and if all the expected hoopla were to accompany such a high figure?

A good aspect of this process is that when a researcher writes down the modifiers, certain undesirable administrative policies will be eliminated because they would be brought formally to the attention of administration, which may not have recognized them as having the weight they did. In essence, the researcher and administration would be forced to face their own objectives and the values assigned to them.

Of course, dishonesty in assigning values is always possible. But then, no system will work in that case.

Some General Problems

It is known that accidents are caused by combinations of multiple factors. When priority of one variable only is evaluated, how is this problem handled?

Validity of the technique for estimating research priority needs to be checked in each of the value-assigning phases, also the reliability.

If a countermeasure is already being implemented, an expression is needed to take account of the fact. It does not exist in the present scheme. It might be a function which can operate on the other implementation functions or weights.

If there is a range of time from beginning to end of implementation, there is a problem of estimating the correct single implementation time, or at least, of dealing with the problem in some way. This is important because in most cases the course of implementation probably does not increase linearly from none to full over the time involved, whereas such linearity is a simplifying assumption in the present formula.

How shall the arbitrary time period for implementation (here, 15 years) be chosen? Is there a rationally derivable value?

What if several countermeasures exist? What if, as is almost always the case, there are several research designs with different costs and different payoffs? How can one get around the necessity of bulky time investment to check out all these possibilities by deriving priorities for each separately?

In cases of in-house research where the people available cannot be increased, and permanent personnel are already paid for, cost figures for them are fixed for all studies. How shall their contribution be counted? This question is important because, conceivably, an optimal set of studies could leave a person idle. What then?

Where does basic research come in? Obviously, some basic research will have value for beyond some applied research. Here basic research is defined as that type of investigation in which the objective is a general increase in the knowledge of the given field. Implied is the absence of a specific goal such as lives saved. The present technique is applicable to basic research only if one can quantify the ways in which one measures the achievement of the stated objectives. On the other hand, it is possible to assign worth to any research in terms of other objectives, whether or not it was originally intended to satisfy those objectives. Keeping to the objective of basic research, no investigation is better than any other, so long as it brings new information. If the statement of objectives is refined to include generalizability of information, or perhaps scope of the dimensions of life that might be affected, or the like, then it becomes easier to quantify the criteria reflecting achievement of the objectives of basic research. By and large, it appears as if the present method would have to be altered considerably, and many simplifying assumptions applied, before it could deal effectively with the question of basic research priorities.

Originally the concept of evaluating priorities took the form of a linear equation. The independent variables were to be factors affecting decisions about priority; the dependent variable was to be a priority score; each variable was to have a weight reflecting its importance in the scheme of priority allocation (e.g., public opinion might have a smaller intrinsic value in an "absolute" sense than that of project cost); and in a given case of research, each independent variable was to be assigned a coefficient reflecting how strongly it applied in the given case. Here, each variable had to be equated to a common measuring unit, presumably also lives or injuries lost or saved. In such an equation, it is obvious that payoff variables are positive and cost variables negative. The independent variables—personnel, time, research effectiveness, inertia of organizational structure, public opinion, policy, etc.—are additive primary variables in this scheme, rather than multiplicative modifiers, as in the scheme developed above.

In considering the factor relating to how much a given research contributes to the total solution, it might be better to assign a value less than one to a project if it only picks up a portion of the total necessary research, even though it may be an essential portion. An alternative way of handling the question may be to split the factor into an importance factor and a proportion-of-effort-required factor. How these would work together has not been examined in detail. At least the problem is recognized.

The present work was not intended in any way to be more than a speculative approach to a problem needing some thinking. There are obviously other approaches (see, for example, the implications of questions and procedures in 4); the technique depends partly on unavailable information; circuitous routes must be taken to arrive at many numbers; refinements are obviously needed for many of the measures; validation will take a long time; the technique is, for some researches, cumbersome; the problem of predicting and evaluating countermeasures effectively is very troublesome; etc.

However, if the paper has led anyone to think about the value system under which he has worked, and to consider the value system under which he may in the future operate, then it will have accomplished one of its own objectives.

REFERENCES

1. "Approaches to an Analysis of the Driving Task." Public Service Research, Inc., Stamford, Conn. (1962). (U.S.P.H.S. Contract SA 76970)
2. Baker, J. S. (Ed.), "Case Studies of Traffic Accidents." The Traffic Inst., Northwestern Univ., Evanston, Ill. (1960).
3. Blotzer, P., Krumm, R. L., Krus, D. M., and Stark, D. E., "Pennsylvania Turnpike Joint Safety Research Group: Accident Causation." Eckhardt, P. K., Ed., Westinghouse Air Brake Co., Swissvale, Pa. (1954).
4. Calica, A. B., Crowther, R. F., and Shumate, R. P., "Enforcement Effect on Traffic Accident Generation." Dept. of Police Admin., Univ. of Indiana, Bloomington (1963).
5. Carpenter, J. A., "Effects of Alcohol on Psychological Processes." Chap. III, "Alcohol and Traffic Safety," Fox, B. H. and Fox, J. H., (Eds.), U. S. Public Health Service, Dept. of H.E.W., Washington, D. C. (1963).
6. Feeley, J. M., Jr., "The Economics of Project Selection in Industrial Research." Res. Rept., Univ. of Chicago School of Business (1954).
7. Fox, B. H., Discussion, "Concept Formation in Accident Research." In "Behavioral Approaches to Accident Research," Assoc. for the Aid of Crippled Children, N. Y. (1961).
8. Goldstein, L., "Accident Prevention Research." Public Health Reports, 78(7), 565-567 (1963).
9. Goldstein, L., "Whither Accident Research?" Traffic Safety Research Review, 7(1), 2-4 (1963).
10. Hemsley, R. T., "The Selection of Research Projects." M.B.A. Thesis, Univ. of Chicago (1953).
11. "Medical Aspects of Motor Vehicle Accident Prevention, a Symposium." N. Y. State J. Med., 56(24), 3851-3882 (1956).
12. Penn, H. S., "Causes and Characteristics of Single Car Accidents: Part One." California Highway Transp. Agency (1963).
13. "Research in Safety Education." Center for Safety Ed., N. Y. Univ., and School and College Conf., Nat. Safety Council (1956).
14. "Research Needs in Traffic Safety." Hearing Before Subcommittee on Traffic Safety, 85th Congress (1958).
15. "Research Needs in Traffic Safety Education." National Commission on Safety Education, National Ed. Assoc., Washington, D. C. (1956).
16. Silver, C., "Performance Criteria—Direct or Indirect." HRB Highway Research Record 55, pp. 54-63 (1964).
17. "Special Report on the Williamsburg Conference: Traffic Safety Research and Human Behavior." In "The Federal Role in Highway Safety." Report to Congress from the Secretary of Commerce (1959).
18. "The Field of Highway Safety Research and the Second Highway Safety Research Correlation Conference (1952)." Nat. Acad. of Sci.-Nat. Res. Council, Publication 454, Washington, D. C. (1956).
19. "Traffic Safety Research Review." Nat. Safety Council, Chicago, quarterly.
20. Weschler, I. R., and Brown, P., (Eds.), "Evaluation Research and Development." Annotated Proceedings, Conf. of Research Administrators, Inst. of Indust. Relations, UCLA (1952).

Performance Criteria—Direct or Indirect

CARL A. SILVER
Franklin Institute Laboratories, Philadelphia, Pa.

•HUMAN FACTORS studies often involve the use of simulation. Simulation permits the economical and easy manipulation of what seem to be the primary variables in a given experimental situation. The general success of simulation studies has led to a growing investment of thought, time and capital in the further development of simulation techniques. However, this investment is producing less than maximal returns due to the increasingly apparent lack of a satisfactory measure of human fatigue. There has certainly been no dearth of attempts to develop such a measure and the very number of these attempts attests to their general lack of success. The implications of a satisfactory measure of fatigue are many. Aside from facilitating evaluation of control systems, information displays, road conditions, etc., a satisfactory measure will act as a common denominator by means of which the separate effects of time on task, task difficulty, loss of sleep, environmental distractions, motivation, etc., on performance potential can be determined. In this paper, from a theoretical consideration of the nature of fatigue, a measure is developed which, hopefully, will fulfill the requirements for a satisfactory measure of fatigue.

PRIMARY CONSIDERATIONS

Driving a car is fatiguing. At least, after driving for many hours, especially in heavy traffic, people will say that they feel tired. One might expect to find these subjective feelings of fatigue reflected in driving performance. That is, one might expect a steady decrement in driving performance corresponding to increasing subjective feelings of fatigue. However, this is not the case. For example, Shaw (35) found that the driver who was fatigued by long periods of continuous driving but was concentrating hard at his task, cannot be distinguished from being fully rested on the basis of what he is doing with his controls or in terms of the actual path traveled by the vehicle, its speed changes, accelerations, etc. (Fig. 1). One explanation of this constancy of performance, first advanced by Krendel and Bloom (27), is that a man adapts to the fatigue or to other stressing conditions by altering his mode of responding; for example, by increasing the amount of effort he puts forth to maintain a constant level of performance. The observation by Shaw that driving performance remains constant during extended periods is not surprising. On the contrary, the observation that performance is unaffected over wide ranges of fatiguing or stressing conditions is very general (1, 18, 28, 38, 39, 40).

If one disregards performance decrements produced by very effortful muscular work such as heavy labor or sprinting, and those decrements produced through extreme environmental conditions such as mechanical vibration of high intensity or extremes of temperature and humidity, the typical form of the performance curve as a function of fatigue is shown in Figure 2. In this figure a generalized form of the curve is shown by the heavy line. Vigilance and tracking tasks show similar curves. It can be seen that, presuming knowledge of results, performance remains essentially constant or decreases only slightly for an extended period of time; subsequently, the performance shows an abrupt decline. It is often expensive or impractical to test a subject for the time interval required in order to show this sharp decrement in performance. It would be desirable to have some measure which would show a steady change related in some simple manner to the amount of work performed, the magnitude of stressing situation, or other relevant variables. Perhaps the sharp decrement in performance which occurs

Paper sponsored by Special Committee on Driving Simulation.

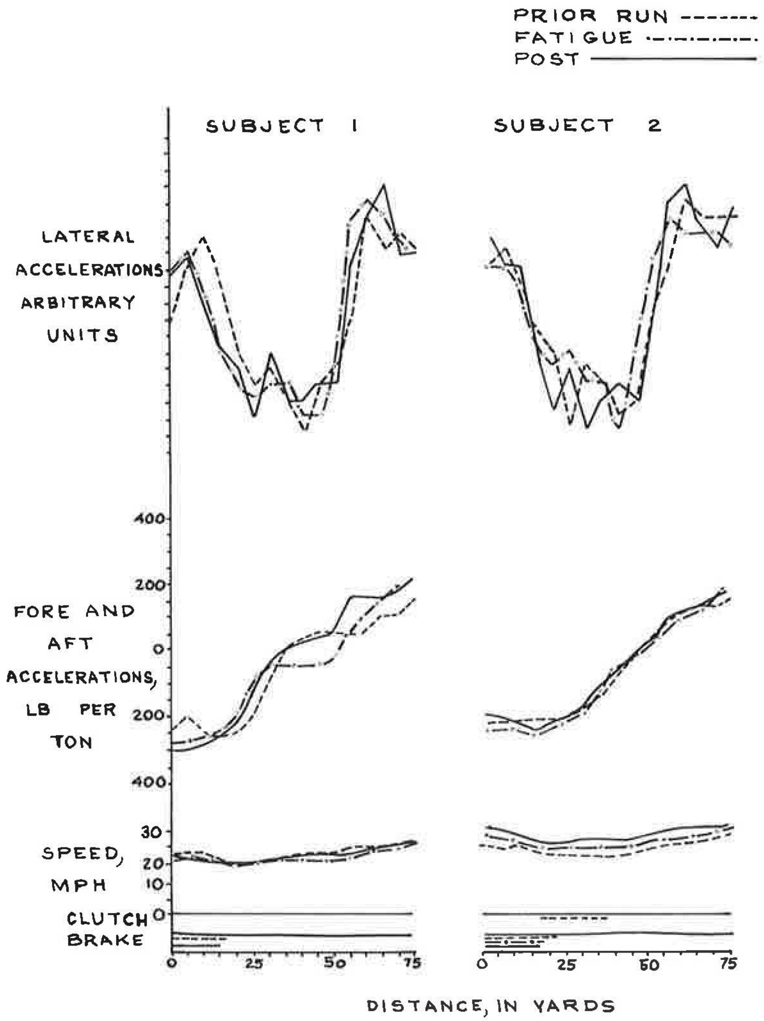


Figure 1.

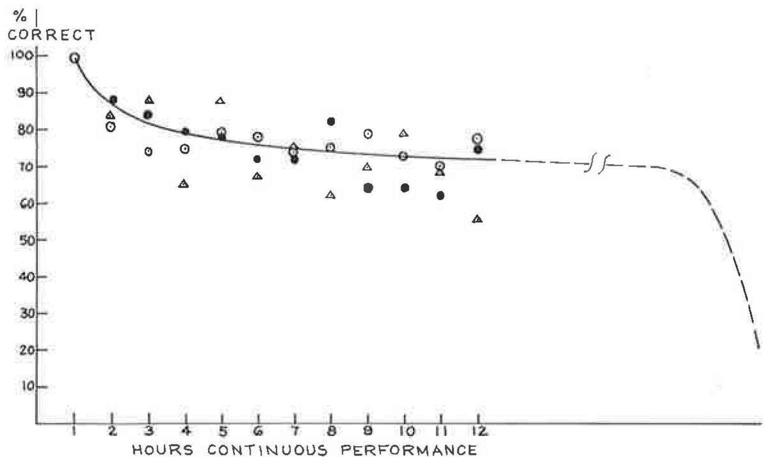


Figure 2. Idealized performance curve. Each point represents averaged hourly performance of four 12-hr sessions by one individual (24).

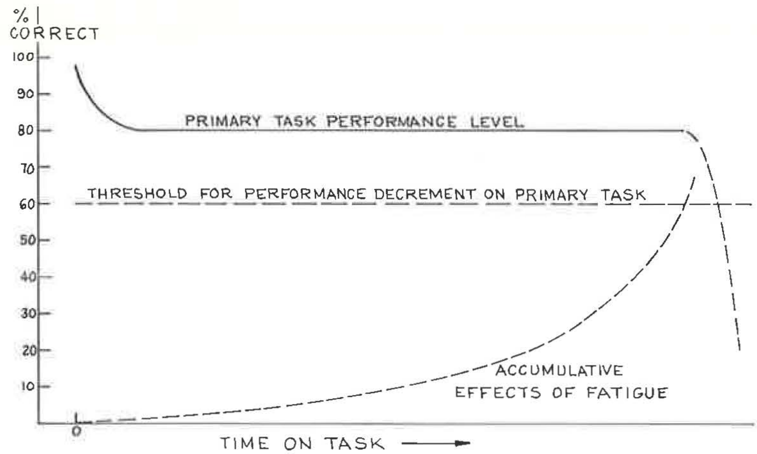


Figure 3. Hypothesized model.

after an extended period of stress or fatigue is not the result of some discrete event occurring at that time but rather is the result of some continuous process which has accumulated to a superthreshold level. Figure 3 shows a theoretical conception of what may be occurring. As before, the heavy line indicates the level of performance, whereas the dotted line shows presumed accumulation of the effects of fatigue, stress, etc. When the accumulated level reaches the indicated threshold, the sudden decrement in performance is observed. A desirable measure of fatigue should show a continuous change and thus allow prediction of when the breakdown will occur as well as comparison of rates of decrement. As previously stated, performance on the criterion task does not show the kind of decrement that is expected.

INDIRECT MEASURES OF FATIGUE

Everyone feels, however, that although objective measures of performance may indicate no significant decrement, the task becomes ever more taxing or arduous with the passage of time. This subjective feeling has led to the search for indirect objective measures of an individual's performance potential; for measures not of the criterion performance but of other behavioral or physiological processes which might be correlated in some way with this performance. One approach is to measure part-processes, i.e., physiological activity, biochemical indicators, and portions of behavioral sequences. Table 1 gives some of these

TABLE 1

INDIRECT MEASURES OF PERFORMANCE

- I. Physiological Indicators
  - A. Physical
    - 1. Respiratory rate
    - 2. Pulse rate
    - 3. Blood pressure
    - 4. Body temperature
  - B. Chemical
    - 1. Contents of blood
      - (a) Glucose
      - (b) Lactic acid
      - (c) Creatine
      - (d) Glutathione
      - (e) Cholinesterase
      - (f) Asorbic acid
    - 2. Contents of urine
      - (a) 17-ketosteroids
      - (b) Potassium ions
      - (c) Calcium ions
      - (d) pH
    - 3. Contents of saliva
    - 4. Metabolic rate
  - C. Electrical
    - 1. EEG
    - 2. EKG
    - 3. EMG
    - 4. GSR
- II. Psychological-Functional Measures
  - A. Sensory
    - 1. Visual
      - (a) CFF
      - (b) Color perception
      - (c) Rate of eye blink
      - (d) Visual acuity
      - (e) Accommodation time
      - (f) Persistence of after images
    - 2. Auditory
      - (a) Absolute thresholds at different frequencies
    - 3. Tactile
      - (a) Absolute pain threshold
      - (b) Absolute vibration threshold
    - 4. Olfactory
      - (a) Absolute thresholds
    - 5. Gustatory
      - (a) Absolute thresholds
  - B. Motor
    - 1. Rate of movement
    - 2. Accuracy of movement
    - 3. Steadiness of movement
    - 4. Muscular strength
    - 5. Muscular endurance
    - 6. Reflex sensitivity

. . . . .and many many more . . . .

TABLE 2

SOME PERFORMANCE MEASURES USED IN RECENT STUDIES

Invest. and Ref.	Measures Used	
Adams et al. (1)	Breathing rate	Skin temperature
	Heart rate	Skin resistance
Frederick (18)	Visual acuity	Kraepelin test
	Dark adaptation	Bourdon-Wiersma test
	Ocular muscular balance	Grunhorn test
	Color perception	Pulse rate
	Absolute auditory threshold	Rate of breathing
	Sense of equilibrium	Blood pressure
	Choice reaction time	17-ketosteroids in the urine
Hussman (23)	Arm steadiness	Critical flicker frequency
	"Blind" arm steadiness	tapping rate
Takakuwa (37)	Critical flicker frequency	Donnagio reaction
	Tapping rate	Ogawa's colloid reaction
	Aesthesiometry	Takano's agglutination test
	Knee reflex	

indirect measures. Most studies of performance decrement or fatigue have sought to establish some correlation between one or more of these measures and performance on the primary criterion task. A review of some typical studies provides an idea of the effectiveness of this approach.

Adams, et al. (1) performed a study in which the results of seven tests of pilot performance and four physiological measurements showed no significant correlation. Frederick (18) made an extensive study of the effects of fatigue resulting from three consecutive days without sleep on the results of the tests given in Table 2. He states, "to our astonishment, the majority of these tests never showed a significant deviation from normal values even during this test of three days." Hussman (23) found that out of four measures only steadiness changes significantly as a function of fatigue. The intercorrelations between the measures were low, however, and the highest correlation, that between steadiness and blind steadiness, which one might expect to approach 1.0, was only 0.55. Takakuwa (37) performed a study on the effect of fatigue resulting from a day's work on each of four functional tests and three biochemical reaction tests. Each of the measures was taken prior and subsequent to the working period. During four years, 333 samples were obtained and the correlations between these tests determined. Statistically significant correlations were obtained between aesthesiometry and critical flicker frequency (-0.524), Takano's agglutination reaction, and critical flicker frequency (-0.287), and aesthesiometry and the knee reflex (0.284). Here, as in Hussman's study, even the highest correlations were relatively low. The correlations between these tests and the criterion performance were not stated, but were evidently low, inasmuch as the authors were trying to validate still another measure.

The results of the preceding four studies are typical in that often no significant effect of fatigue or environmental stress on the various measures was noted, and where there were significant changes in some measures, their intercorrelations and/or their correlations with the criterion performance were low. In light of these results, the random selection and evaluation of measures which will hopefully be sensitive to changes in the subject's level of fatigue are inefficient and unproductive. Perhaps fuller theoretical consideration of the nature of the processes underlying performance decrement will provide the necessary rationale by means of which successful measures of performance decrement may be developed.

SUBSIDIARY TASKS

To provide a groundwork for such a theory, another approach to prediction of performance decrement is examined. Thus far concern has been with measures arising from functions of only a part of the organism. Except for measures of the criterion task itself, there has been concern with discrete physiological or behavioral processes. Another interesting approach to the prediction of performance decrement is to measure how well a man performs a second task simultaneously with the criterion activity. This is often known as the "subsidiary workload" method. This technique resulted from earlier work on the division of attention and has been used by Bornemann (6, 7), to study the interactions between a subsidiary task of mental addition and a wide range of primary tasks. It has since been used to study automatization of performance resulting from practice; Bahrick et al. (4), Bahrick and Shelley (3), Broadbent (8), to measure the relative difficulty of listening tasks, Broadbent (9), to measure the order of difficulty of motor tasks, Poulton (31), and to compare man-machine tracking systems, Garvey and Taylor (22). Poulton, following Broadbent, stated the basis of this technique as "the assumption that there is a limit to the rate at which an operator can deal with information—in other words he has a limited 'channel capacity.' When the demands of the primary and subsidiary task together exceed this limit, errors must occur."

The subsidiary workload method is of value in measuring task difficulty, at least in the sense of the magnitude of information transduction required. This method has also been applied to the measurement of the effects of fatigue, but with less than ideal success. Although it is undoubtedly felt that a task is more "difficult" when a man is fatigued, this is probably not true in terms of the amount of information that is processed.

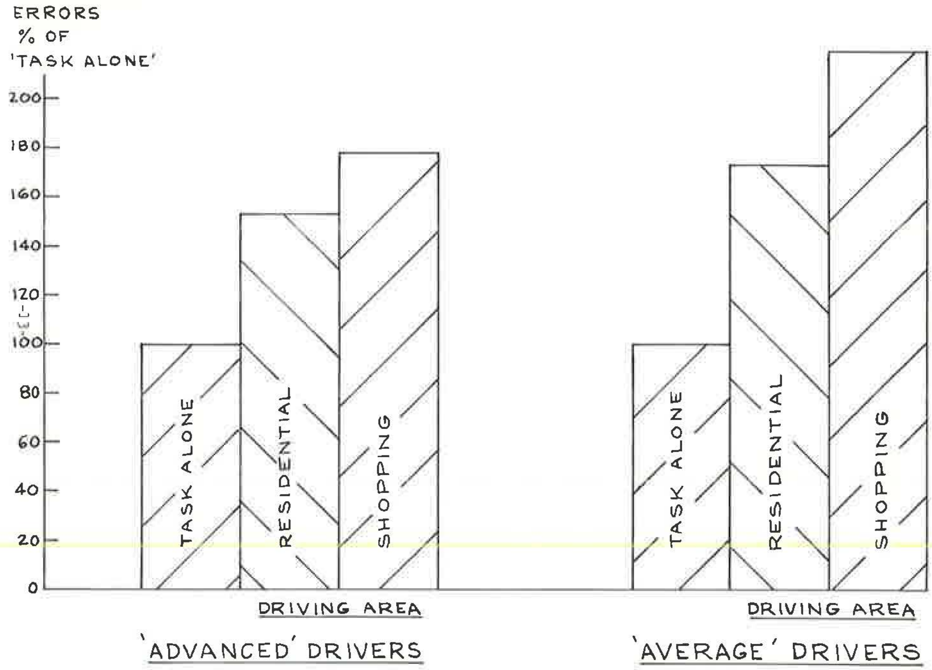


Figure 4. Mean errors on subsidiary task as a function of primary task difficulty (10).

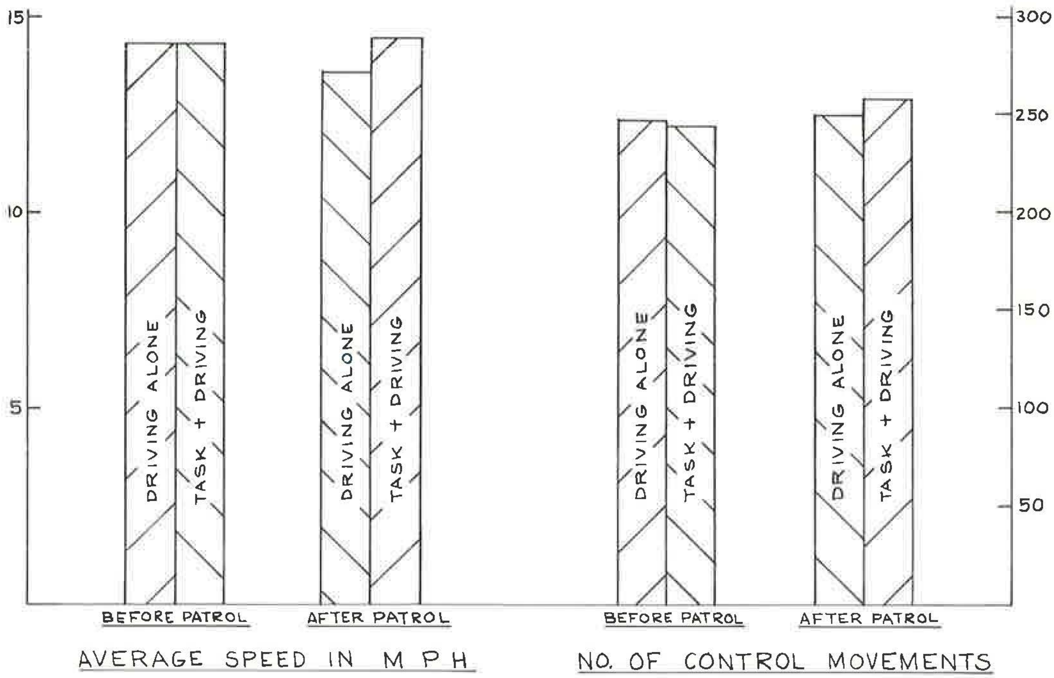


Figure 5. Driving performance as a function of fatigue and subsidiary task (11).

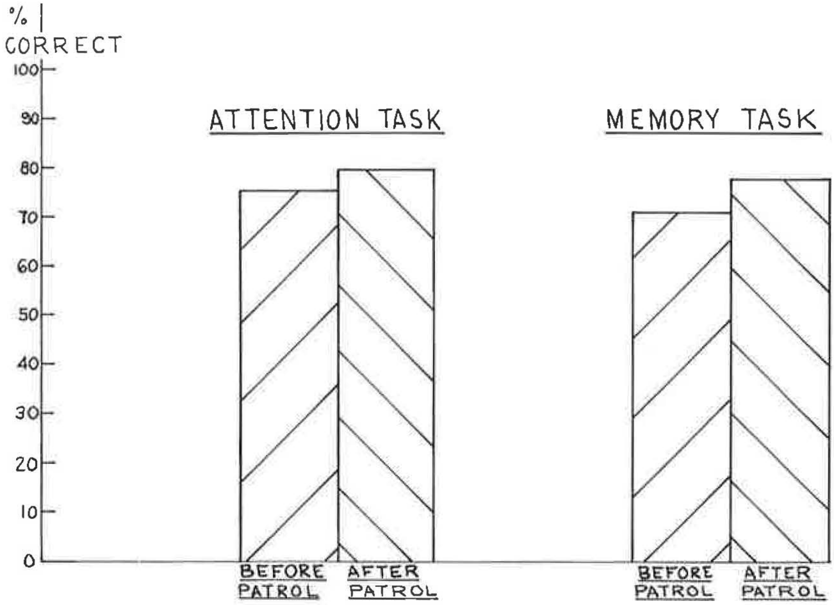


Figure 6. Subsidiary task performance as a function of fatigue (11).

SOME UNEXPECTED DATA

Figure 4 shows that subsidiary task performance nicely reflects the difficulty of the primary task for both "advanced" (N=8) and "average" (N=7) drivers. Performance on the subsidiary task is poorest while driving in a shopping area, better while driving in a residential area, and best while performing the subsidiary task alone. Having seen that task difficulty affects subsidiary task performance, the question arises as to whether the addition of a subsidiary task affects primary task performance. Brown (11) has provided an answer. Figure 5 shows that there is no significant effect of a similar subsidiary task on driving performance as measured by average speed and number of control movements either before or after an 8-hr driving patrol. Thus, subsidiary tasks are sensitive to the difficulty of the primary task, but the primary task performance is relatively insensitive to the addition of a subsidiary task. Can the subsidiary task also be used to measure fatigue? Brown (11) asked this question and received the answer shown in Figure 6. There is a definite effect of fatigue on subsidiary task performance but in the opposite direction to what one would expect. This figure shows that the number of correct responses increased as a result of 8 hours driving on patrol. Brown gave three possible, circumstantial explanations for this finding. Perhaps there is another, more basic, explanation for this initially disturbing result.

AN EXPLANATION

Here is an apparently typical picture. The primary task performance has remained constant over a relatively extended period and the subsidiary task not only does not show a decrement, but actually shows an increment. Probably there is a relatively simple explanation of what has occurred that gives an important clue as to how a really useful measure of performance decrement can be achieved. First, the reasonable notion should be accepted that the drivers did indeed become fatigued as a result of 8 hr driving in a patrol car. Second, the increase in performance of the subsidiary task should be accepted as genuine. It is noted that, as has been repeatedly demonstrated, the onset of fatigue is characterized by an increase in muscular tension (Freeman, 19). An increase in tenseness (Jacobson, 26) is a facilitator for many responses ranging from simple reaction time to mental arithmetic. The classic example of this effect is the facilitation of mental arithmetic and other responses resulting from squeezing a hand dynamometer while performing these tasks. Therefore, assuming that the drivers did become fatigued as a result of their 8-hr patrol and therefore were somewhat more tense at the end of the 8-hr period than they were at the beginning, it might be predicted that their performance on this auditory task would be increased.

If this explanation is correct, then the use of subsidiary workload as a measure of fatigue is not theoretically justified. A distinction must be made between the effects of increasing task difficulty and the changes which result from fatigue. In the case of task difficulty the concept of channel capacity is not unreasonable. Bahrick et al. (4) found, for example, that errors in subsidiary task decreased as subject acquired skill on a primary task. It is reasonable to suppose that at least part of the skill involved in learning the primary task was that of not attending to irrelevant cues. Thus, as the subject became more skilled, he processed less irrelevant information, using less "channel capacity," and thus making more available for the subsidiary task, the result being a decrease in the number of errors in the subsidiary task as the subject learned the primary task. The concept of channel capacity is essentially a "central" concept. That is, it refers to a process which affects all tasks in about the same way. This is not true with regard to tension as has been conceived. Tension is a peripheral phenomenon.

The fatigue of responses is relatively specific to the particular response elements involved. The interaction of tasks or responses as one fatigues, is a function of distance from the response element which is being fatigued to the point of measurement. Therefore, the degree to which a subsidiary response will be affected by the fatiguing of a primary response is a function of the spatial relations between the respective response elements involved. The evidence for such facilitative interaction is clear (Freeman, 20), (Jacobson, 25). Apparently the locus of the responses, which are called "mental activity," is rather diffuse. At least it seems that practically any increase in body tension can have a facilitative effect on "mental" tasks. If "mental" or perceptual responses are relatively diffuse it is easy to appreciate how the fatigue of driving, resulting in increased tension distributed widely over the body, could interact to facilitate mental and perceptual tasks, thus accounting for Brown's data. However, the effect of increasing tension is not always facilitory. Stauffacher (36) and Courts (12) have shown that there is first an increase and then a decrease in performance as tension increases. The reason for the interference is apparently somewhat as follows: When only a moderate amount of tension is induced there is facilitation of other simultaneously occurring responses. These responses, however, can only be facilitated up to some physiological limit. Beyond this limit there is no further facilitation; but responses, often competing responses, of lower habit strength are likewise facilitated. The result is that the differential in response strength between the "correct" response and competing response is decreased. Thus there is an increased response variability coupled with a decrease in the probability of a correct response. When tension is increased, ultimately a point is reached where more inappropriate facilitation occurs than appropriate, and from that point on increasing tension appears to be inhibitory rather than facilitory in terms of the correctness of the responses being made. Thus, early in learning, increased tension may be facilitory because the correct responses are not near their physiological limit, whereas late in learning, increasing tension tends to be inhibitory because the competing "incorrect" responses are facilitated to a greater extent than the "correct" responses. In this way, it is found that "mental blocking" is associated with fatigue (Ash, 2), (Bills, 5). When competing responses are facilitated to the point where they are approximately equal in strength to the correct response tendencies, there is an internal "dilemma." Response time may be then greatly increased.

In a sense, then, fatigue is the opposite of learning. During learning, extraneous facilitative responses are eliminated. Because interaction is a function of proximity, the sequence of elimination is important. Renshaw and Schwarzbek (32) took movies of a subject practicing a tracking task. They found that the trunk muscles relaxed first, then the shoulders, upper arm, and lower arm, in that order. More distant muscle groups are also involved in the changes. Freeman (19) found a progressive reduction in leg muscle-tension as the subject learned either a manual tracking task or mental arithmetic. During learning the direction of change is from distribution to focalization of muscular activity. During fatigue, however, the reverse of this process occurs even though performance measures remain constant. Muscular tension is generalized under most of the conditions that are associated with effortful performance. This was first pointed out by Duffy (15) and has been studied by Ryan, Cottrell, and Bitterman (34). Factors thus far shown to increase the spread of muscular tenseness are lack of practice; prolonged work periods (Ash, 2), (Robinson and Bills, 33); distraction (Morgan, 30), (Davis, 13); increased difficulty of task (Davis, 14), (Eason and White, 16); and increased incentives (Freeman, 21). This formulation of the effects of muscular tension is adapted from Meyer (29) and from earlier unpublished work by the same author. The responsibility for errors in interpretation or application of this formulation, however, must lie with the present author.

A POSSIBLE MEASURE

On the basis of the foregoing discussion, an indirect measure for predicting response decrement is examined. This measure is the distribution of muscular activity during a behavioral "instant" (a time interval short enough so that sequential responses will not occur); all muscular activity measured during that instant may be taken as being simultaneous. One method of measurement is as follows: a suitably large number (e.g., 20) of electrodes are placed upon a subject and are distributed widely over the body surface. The resting level of muscular activity at each electrode is noted. Then, as the subject performs the fatiguing task (e.g., driving a car or a simulator) the number of electrodes exhibiting activity above the initial level are determined during a behavioral instant. Successive measurements during the course of fatiguing activity should exhibit an increase in the number of electrodes showing increased activity, thus indicating an increase in the distribution of tension over the body. As recruitment proceeds and the distribution of response rises, there will come a point at which the recruiting of additional response elements occurs at so great a distance from the fatiguing response site that the contribution of the newly recruited units is not sufficient to maintain the required performance level. At this point, as shown in Figure 1, a sharp decrement in performance will be observed.

Theoretical justification alone, however, indicates nothing about the feasibility of the measure. A measure should be (a) within the technological state of the art; (b) immediately available (not requiring processing, developing, etc.); (c) not damaging to the subject; and (d) relatively economical. Within the past decade, technological progress in bioelectronics has been such that development of equipment for the amplification, integration and recording of muscle potentials entails merely the selection of parameters to optimize the equipment for use in any specific application. An excellent survey of bioelectronics is provided by Ford (17).

Inasmuch as the integrations will be performed within a "behavioral instant" (approximately $1/10$ of a sec), the values of all 20 points will be available for recording 10 times per sec, thus satisfying the criterion of "immediate" availability. Although surface electrodes will be used, which will restrict the subjects range of movement, they will cause no damage, and little discomfort to the subject. Finally, the development of high-performance transistors and circuitry permit the design of relatively economical circuits using a minimum number of components to achieve maximum effectiveness of operation. It is hoped that this new approach will provide a reliable, indirect measure of the rate of recruitment by means of which the point of sharp performance decrement can be predicted.

REFERENCES

1. Adams, O. S., Levine, R. B., and Chiles, W. D., "Research to Investigate Factors Affecting Multiple Task Psychomotor Performance." Lockheed Aircraft Corp., USAF WADC Tech. Rept., (Mar), No. 59720, VII, p. 37 (1959).

2. Ash, I. E., "Fatigue and Its Effect upon Control." Arch. Psychol., 31, pp. 1-61 (1914).
3. Bahrick, H. P., and Shelley, C., "Time-Sharing as an Index of Automatization." J. Exp. Psychol., 56, pp. 388-393 (1958).
4. Bahrick, H. P., Noble, M., and Fitts, P. M., "Extra-Task Performance as a Measure of Learning a Primary Task." J. Exp. Psychol., 48, pp. 298-302 (1954).
5. Bills, A. G., "Fatigue, Oscillation, and Blocks." J. Exp. Psychol., 18, pp. 569-570 (1935).
6. Bornemann, E., "Untersuchungen uber der Grad der Geistigen Beanspruchung. I. Aus Arbeitung der Methode." Arbeitsphysiologie, 12, pp. 142-172 (1942).
7. Bornemann, E., "Untersuchung uber der Grad der Geistigen Beanspruchung. II. Praktische Ergebnisse." Arbeitsphysiologie, 12, pp. 173-191 (1942).
8. Broadbent, D. E., "Listening Between and During Practiced Auditory Distractions. Brit. J. Psychol., 47, pp. 51-60 (1956).
9. Broadbent, D. E., "The Bass-Cutting of Frequency-Transposed Speech." M.R.C. Appl. Psychol. Unit Rept. No. APU 223 (1956).
10. Brown, I. D., and Poulton, E. C., "Measuring the Spare Mental Capacity of Car Drivers by a Subsidiary Task." Ergonomics, 4:(1) 35-40 (1961).
11. Brown, I. D., "Measuring the Spare Mental Capacity of Car Drivers by a Subsidiary Auditory Task." Ergonomics, 5:(1) 247-250 (1962).
12. Courts, F. A., "Relations Between Experimentally Induced Muscular Tension and Memorization." J. Exp. Psychol., 25, pp. 235-256 (1939).
13. Davis, R. C., "The Relation of Certain Muscle Action Potentials to 'mental work'." Ind. Univ. Publ. Sci. Ser., No. 5 (1937).
14. Davis, R. C., "The Relation of Muscle Action Potentials to Difficulty and Frustration." J. Exp. Psychol., 23, pp. 141-158 (1938).
15. Duffy, E., "The Relation Between Muscular Tension and Quality of Performance." Amer. J. Psychol., 44, pp. 535-546 (1932).
16. Eason, R. G., and White, C. T., "Muscular Tension, Effort and Tracking Difficulty: Studies of Parameters Which Affect Tension Level and Performance Efficiency." Perceptual and Motor Skills, 12, pp. 331-372—Monograph Supplement 4-V12 (1961).
17. Ford, A., "Foundations of Bioelectronics for Human Engineering." Navy Electronics Lab., Rept. 761, p. 119 (1957).
18. Frederick, W. S., "Physiological Aspects of Human Fatigue." A.M.A. Arch. Ind. Health, 20, pp. 297-302 (1959).
19. Freeman, G. L., "The Spread of Neuro-Muscular Activity During Mental Work." J. Gen. Psychol., 5, pp. 479-494 (1931).
20. Freeman, G. L., "Mental Activity and the Muscular Processes." Psychol. Rev., 38, pp. 428-447 (1931).
21. Freeman, G. L., "The Facilitative and Inhibitory Effects of Muscular Tension upon Performance." Amer. J. Psychol., 45, pp. 17-52 (1933).
22. Garvey, W. D., and Taylor, F. V., "Interactions Among Operator Variables, System Dynamics and Task-Induced Stress." J. Appl. Psychol., 43, pp. 79-85 (1959).
23. Hussman, T. A., "Indicators of Behavior Decrement: A Critical Evaluation of Four Indicators of Behavior Decrement." Dept. of Army, DA-49-007-MD-222 (O.I. 19-52), Tech. Rept. No. 12, #31301 (1952).
24. Huxtable, Z. L., White, M. H., and McCartor, M. A., "A Re-performance and Re-interpretation of the Arai Experiment in Mental Fatigue with Three Subjects." Psychol. Monogr., 59 (1946).
25. Jacobson, E., "Progressive Relaxation." Univ. Chicago Press. (1938).
26. Jacobson, E., "Muscular Tension and the Estimation of Effort." Amer. J. Psychol., 64, pp. 112-116 (1951).
27. Krendel, E. S., and Bloom, J. N., "The Natural Pilot, A Criterion System for Flight Proficiency Evaluation." Franklin Inst. Rept. F-A2151-1, (Sept. 1959) (prepared for U. S. Naval Training Device Center).

28. Mackworth, N. H., "Researches on the Measurement of Human Performance." Medical Research Council, Spec. Rept. No. 268 (London, H.M.S.O.) (1950).
29. Meyer, D. R., "On the Interaction of Simultaneous Responses." Psychol. Bull., 50, pp. 204-220 (1953).
30. Morgan, J. J. B., "The Overcoming of Distraction and Other Resistances." Arch. Psychol., 35 (1915).
31. Poulton, E. C., "Measuring the Order of Difficulty of Visual-Motor Tasks." Ergonomics, 1, pp. 234-239 (1958).
32. Renshaw, S., and Schwarzbek, W. C., "The Dependence of the Form of the Pursuit-Meter Learning Function on the Length of the Inter-Practice Rests: I. Experimental." J. Gen. Psychol., 18, pp. 3-16 (1938).
33. Robinson, E. S., and Bills, A. G., "Two Factors in the Work Decrement." J. Exp. Psychol., 9, pp. 415-443 (1926).
34. Ryan, T. A., Cottrell, C. L., and Bitterman, M. E., "Muscular Tension as an Index of Effort: The Effect of Glare and Other Disturbances in Visual Work." Amer. J. Psychol., 63, pp. 317-341 (1951).
35. Shaw, W. J., "Objective Measurement of Driving Skill." International Road Safety and Traffic Review, pp. 21-30 (1957).
36. Stauffacher, J. C., "The Effect of Induced Muscular Tension upon Various Phases of the Learning Process." J. Exp. Psychol., 21, pp. 26-46 (1937).
37. Takakuwa, E., "The Function of Concentration Maintenance (TAF) as an Evaluation of Fatigue." Ergonomics, 1, pp. 37-49 (1962).
38. Thorndike, E. L., "Mental Work and Fatigue and Individual Differences and Their Causes." Educational Psychology, 3:69, N. Y. Teach. Coll., Columbia Univ. (1914).
39. Thorndike, E. L., "The Curve of Work and the Curve of Satisfyingness." J. Appl. Psychol., 1, pp. 265-267 (1917).
40. Welford, A. T., "The Psychologists Problem in Measuring Fatigue." In "Symposium on Fatigue," W. F. Floyd and A. T. Welford, Eds., London, H. K. Lewis for the Ergonomics Research Society (1953).