

THE EFFECT OF VISUAL TASK STIMULUS-RESPONSE COMPATIBILITY ON SIMULTANEOUS AUDITORY TASK PERFORMANCE

Thesis for the Degree of M. A.
MICHIGAN STATE UNIVERSITY
REID P. JOYCE
1968

THESIS

LIBRA(C)
Michael n See

ABSTRACT

THE EFFECT OF VISUAL TASK STIMULUS-RESPONSE COMPATIBILITY ON SIMULTANEOUS AUDITORY TASK PERFORMANCE

by Reid P. Joyce

This study attempted to demonstrate that stimulusresponse compatibility in a forced-pace sequential visualmotor task can be examined using a technique previously
employed mainly to study operator workload. S-R compatibility has in the past been studied in non-paced
reaction time situations. Workload has been studied
through the use of subsidiary task measures. It was
hypothesized here that S-R compatibility is a contributor
to operator workload, and therefore changes in S-R compatibility can be examined indirectly by the subsidiary
task technique of measuring workload.

Four experimental conditions were used, in which four different S-R configurations, representing different degrees of S-R compatibility, were imposed on a visual-motor task. The relation between patterns of extinguished lights in a matrix and the horizontal positions of four pushbuttons used to relight the lights was varied. The relations were: 1) direct spatial; 2) reverse spatial; 3) symbolic (numerical); and 4) random.

The simultaneously performed secondary task, whose error rate was to be used as an index of difficulty of

the primary task, was an auditory inspection task in which the subject heard (through headphones) a tape-recorded series of numbers and made verbal responses to new digits appearing in the series.

The secondary auditory task error rate was found to be sensitive (p <.01) to changes in the primary task S-R configurations. However, an interaction was revealed between the numerical mediator used in the symbolic condition of the visual task and the numerical verbal responses to the auditory task. This interaction took the form of visual (numerical) responses intruding into the auditory task responses, but not vice-versa. The validity of the rank order of S-R compatibility as defined by this particular secondary task error rate was therefore questioned, because it was felt that the nature of its interaction with the primary task disqualified this secondary task as a good tool for measuring workload with this particular primary task.

Approved:

Committee Chairman

Date:

Thesis Committee:

Dr. Theodore W. Forbes, Chairman

Dr. Paul Bakan

Dr. Gordon Wood

THE EFFECT OF

VISUAL TASK STIMULUS-RESPONSE COMPATIBILITY ON SIMULTANEOUS AUDITORY TASK PERFORMANCE

By

Reid P. Joyce

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

951311

ACKNOWLEDGMENTS

It is difficult to convey in a short section such as this the great outpouring of gratitude and relief that one feels a need to express upon completion of a thesis. impact of various people and events on the course of a project such as this is much clearer now than it was when the opportunity existed for a real, first class, face-toface, emotional expression of thanks in appropriate amount and kind. My wife Beth occupies the top position in the list of contributors, but I'll thank her personally rather than try to enumerate her contributions here. It turns out that everyone else is tied for second position in the list, so the following is to be considered a nominal rather than ordinal arrangement: Dr. T. W. Forbes, committee chairman, mentor, and employer, was helpful, thoughtprovoking, and very patient, and I am sure I will carry his influence far beyond this thesis; Rick Pain and John Fry contributed some good thoughts and prevented some bad ones from being included, and they helped immeasurably by absorbing some of my non-thesis workload during the running of the experiment; Doug Williams made substantial contributions to the design and construction of the equipment, and to my understanding and enjoyment of Baroque music.

TABLE OF CONTENTS

Secti	Lon No.	Page No.
I.	INTRODUCTION	1
II.	METHOD	15
III.	RESULTS	26
IV.	DISCUSSION AND SUMMARY	31
Refer	cences	47

LIST OF TABLES

Table	No.		Page No.
1.	Group Mean	Differences and Critical	
	Difference	for Subject Mean Auditory	
	Errors Per	Trial	28
2.	Group Mean	Differences and Critical	
	Difference	for Subject Mean Visual	
	Errors Per	Trial	33

LIST OF FIGURES

Figure	No.	Page	No
1.	Stimulus Configurations Corresponding to		
	Response Button Positions for Four Stimulus		
	Conditions	18	3
2.	Effect of S-R Compatibility on Mean Auditory	7	
	Task Errors Per Trial	2	7
3.	Effect of Trials on Mean Auditory Task		
	Errors	30)
4.	Effect of S-R Compatibility on Mean Visual		
	Task Errors Per Trial	3:	2
5.	Effect of Trials on Mean Visual Task Errors	3.	5

LIST OF APPENDICIES

Appendix	No.	Page No.	Page No.				
1.	Instructions	to	Subjects	43			

SECTION I

INTRODUCTION

This study attempted to demonstrate that stimulusresponse compatibility in a forced-pace sequential visualmotor task can be examined using a technique previously
employed mainly to study operator workload. S-R compatibility has in the past been studied in non-paced
reaction time situations. Workload has been studied
through the use of subsidiary task measures. It was
hypothesized here that S-R compatibility is a contributor
to operator workload, and therefore changes in S-R compatibility can be examined indirectly by the subsidiary
task technique of measuring workload.

Background

Operator Workload. In the design of equipment and tasks for human operators, the situation often arises in which an operator must be able to perform two or more tasks concurrently. There is presumably a limit to the amount of workload that a given operator can handle, so it is important, for a given task, that something be known about the extent to which that task can be shared with other tasks. If the task keeps the operator so busy (physically and/or mentally) that he cannot adequately perform other tasks, then it would be foolish to assign other tasks to him. By the same token, it may be an advantage to know whether an apparently difficult job is

SECTION I

INTRODUCTION

Fig study attempted to demonstrate that stimulusresponds consatibility in a forced-pace sequential visualmotor task can be examined using a technique previously
employed mainly to study operator workload. G-R compatibility has in the past been studied in non-paced
reaction time situations. Morkload has been studied
through the use of subsidiary task measures. It was
hypothesized has either Generalibility is a contributor
to operator workload, and therefore changes in G-C compatibility can be examined inth etty by the subsidiary
task technique of measuring workload.

निवार ए बहुत कृत्य

tacks for beson operators, the situation of equipment and tacks for beson operators, the situation often arises in which an operator must be able to perform two or more tacks concurrently. There is areasenably a limit to the amount of verbload that a given operator can tan'le, so it is important, for a given tack, that something be brown and the catent to which that lock each to be with other tacks. If the tack beens the epotetic simple with other tacks, if the tack beens the epotetic simple to the performance of a tack of the entire tack of the epotetic simple to a second of the factor of the second of the factor of the each of the early performance of the entire tacks to the second of the factor of the each of the factor of the each of the each of the entire tacks to the entire tacks the entire tacks to the entire tacks to the entire tacks to the entir

really occupying all of the operator's capabilities. Sometimes a relatively simple task requires so much motor activity that the operator could not possibly perform another task concurrently -- he might be able to process more information, but would not be able to make appropriate responses. On the other hand, a task requiring relatively little overt activity might place such a load on the operator's information processing capabilities that he could not respond to additional stimuli even though (in the sense of having his hands free) physically able to do so. Operator workload or operator loading, then, as the term will be used here, refers to the extent to which performance of a particular task precludes or limits other concurrent activities. The source of this interference, and the nature of the "other" activities in question will always be key factors in determining operator load for a given task.

Although the present study was not designed as an information theory study, some of the concepts used in information theory will be useful in making explicit some of the assumptions upon which the present study was based. Operator workload, when the term is used in the information processing sense, may be thought of as the proportion of an individual's channel capacity which is devoted to the performance of the task in question. When the individual is subjected to multiple-sensory

inputs (or multiple inputs in one sense mode) he may be able to apply a coding operation and thereby process the inputs simultaneously (if the inputs can be related or coded in some meaningful way). Or he may have to time-share the inputs, alternately sampling and processing first one input and then the other. As will be pointed out later, the present study assumes that if two tasks are selected which involve unrelated kinds of information, which cannot be meaningfully coded, and if inputs for the tasks are presented to two different sense modes, the tasks will have to be performed on a time-shared basis.

Measurement of Workload. Direct performance measures are often not sufficient to reflect differences in difficulty of various tasks, any of which can be performed adequately. If each task can be performed to its respective criterion, measures of overt responses will not necessarily provide a basis for comparing compatibility of the tasks with other concurrent activities. That is, these measures do not indicate the extent to which a particular task loads an operator's capacity for performing concurrent activities.

A number of investigators have attempted to get at the problem of operator loading by using indirect measures, in the form of subsidiary or secondary task performance. Performance on a particular secondary task, performed concurrently with the primary tasks, is compared for the different primary tasks under investigation. If performance on the different primary tasks remains relatively unaffected by the addition of the secondary task, then differences in secondary task error rates are attributed to differences in the operator loading requirements of the primary tasks. That is, a higher error rate on the secondary task indicates greater difficulty or operator loading requirements associated with the primary task.

A variety of secondary tasks have been utilized by a number of investigators. Bahrick, Noble, and Fitts (1954) used a mental arithmetic task to measure learning on a visual-motor task which required subjects to push a button whenever a light on a moving drum passed under a reference line. By introducing the secondary task either early or late in practice of the visual-motor task the investigators were able to demonstrate that subjects who received the visual stimuli in repetitive patterns performed the mental arithmetic significantly better after visual task practice than did subjects who had had the same amount of practice on the visual task, but with random stimulus presentations. Bahrick and Shelley (1958) used performance on an auditory-motor task as an index of "automatization" in a visual-motor task. As the primary task, subjects pushed one of four buttons when one of four lights came on. There was a straight horizontal relation between the buttons and the lights. As the secondary task, they pushed one of five buttons when they heard the numbers

1 through 5 at irregular intervals. Four visual task conditions ranged from repetitive (in which the pattern repeated after every fourth presentation: 1,3,2,4,1,3,2,4, etc.) through two degrees of redundancy (frequent occurrences of selected diagrams: 1-3, 3-2, 2-4, 4-1) to a completely random sequence. They found that the auditory task interfered with the visual task, but less so as "automatization" (redundancy) increased.

Benson, Huddleston, and Rolfe used a number of peripheral measures to evaluate two aircraft altimeter displays used in a tracking task. In addition to a number of physiological measures, they used a secondary task in which subjects pushed one of two buttons in response to the presence of one of two lights. In the absence of the secondary task, the physiological measures (heart and respiration rate, skin resistance, etc.) failed to discriminate between the two displays' difficulty of use. When the secondary task was added to the tracking task, not only did secondary task error rate suggest (p = .001) that one display was easier to use than the other, but the additional workload of the combined tasks brought difficulty to a level at which the combined physiological measures were able to discriminate (p = .005) between the two displays (physiological measures and secondary task both recommended the same display).

Brown (1962) attempted to measure fatigue in policemen completing eight-hour driving shifts by using two

different subsidiary auditory tasks: 1) a continuous series of random digits which the subject "searched" for three consecutive digits in the order "odd-even-odd": and 2) a series of ten letters, one letter every five seconds, with two identical and eight different letters. The subject reported which letter occurred twice. found small, non-significant before-driving/after-driving differences in subsidiary task performance, but had some difficulty explaining the fact that most subjects performed better after a driving shift than before it. Brown and Poulton (1961) used a subsidiary task similar to that described by Poulton (1960) in an attempt to compare the difficulty of driving in "residential" areas with that of driving in "shopping" areas. The task to be performed concurrently with driving was an auditory inspection task in which the subject listened to a long series of eightdigit numbers, each number differing from the preceding number in only one digit. To make a correct response, the subject said aloud the new digit in each number in the series. As predicted, the subjects produced higher subsidiary task error rates in the "shopping" areas. A second group of drivers, using mental addition of groups of three digits as a secondary task, yielded similar results.

Garvey and Taylor (1959) used several loading tasks (e.g., mental arithmetic, detecting and reporting range and bearing of simulated radar targets) in a somewhat

different way from the studies mentioned above. The investigators added the loading tasks to two different tracking systems in order to observe deterioration in the tracking (primary) task performance. They found that the tracking system which had been found to be superior in an unstressed (no loading task) situation remained superior when loading tasks were added, i.e., performance decrement for the easier system was less than that for the difficult system. Most of the other studies cited above tried to use subsidiary tasks which produced little or no performance-degrading interference with the primary tasks.

Knowles (1963), in a discussion of desirable characteristics of loading tasks, mentioned the following characteristics:

- 1. Non-interference. If the secondary task performance is to be a measure of the load imposed by the primary task, then the secondary task should not physically interrupt or otherwise interfere with the primary task.
- 2. Simplicity. "Ideally, the task should require very little learning and should show little inter-subject variability."
- 3. <u>Self-Pacing</u>. Self-adjusting automatic feedback systems, which adjust loading task presentation rate as a function of operator performance, are recommended.

- 4. Scoring. "The index of operator-load that is calculated from the scores of a given loading task should be comparable from situation to situation."
- 5. <u>Compatibility (Intertask</u>). The loading task should be different from the primary task, and, if possible, it should simulate the kinds of concurrent activities which may be required of the operator in addition to the primary task.

Knowles also cautions that we be aware of the nature and limitations of such measures of operator load:

It is well to look more closely at what may be expected of any measure of operator work-load derived from auxiliary task performance. Fundamentally, such measures yield an ordinal scale; 100 per cent auxiliary task performance does not mean zero operator loading, nor does zero auxiliary task performance mean 100 per cent operator loading. Furthermore, equal increments in loading scores most certainly do not reflect equal increments in work-load. It is therefore most prudent to regard whatever numerical values that are derived with some modesty and to call them what they are -- simply indices of operator-load.

To summarize, then: secondary tasks have been used successfully to derive indices of operator loading requirements for various primary tasks. These indices generally result in an ordinal ranking of alternative forms of the primary task. When used in this way, secondary tasks should disrupt the performance of the primary task as little as possible; they should require little learning; they should produce relatively little intersubject variability; they should, if possible, be different from the primary tasks with which they are to be

used; and they should, if possible, simulate other tasks which may have to be performed with the primary task, if the primary task is to be used in a real system.

S-R Compatibility. Several investigators have discussed the concept of display-control or stimulusresponse compatibility. This is generally defined as the degree to which controls seem to "go with their related display, or the degree of "naturalness" of the relationship. It has been noted (Ross, Shepp, and Andrews, 1955, and McCormick, 1964) that there are population stereotypes with regard to preferences of certain kinds of relationships in certain situations. These preferences are usually culturally based, and have typically been reinforced many times in the individual's history (e.g., light switch on the wall goes up to turn on (in the United States, but down in some other countries); cold water is on the right; bottom elevator button means "going down"). S-R configurations which violate these expections are considered incompatible, and have been found to be more difficult to learn and to perform than those which conform. difference seems to be stable over time, despite continued practice.

Fitts and Deininger (1954) presented subjects with a number of different stimulus dimensions to which they were to respond by moving a stylus from the center to one of eight positions on the circumference of a circular display, the positions corresponding to 12 o'clock, 1:30, 3 o'clock,

4:30, 6 o'clock, 7:30, 9 o'clock, and 10:30. The stimulus dimensions were: 1) two dimensional, spatial; consisting of a circular arrangement of eight lights; 2) one dimensional spatial: a row of eight stimulus lights; 3) two dimensional symbolic: a window with a clock time appearing in it (e.g., 9 o'clock) and 4) non-spatial symbolic, with a window with a non-clock-related word appearing in it (e.g., proper name Joe). Several degrees of correspondance among elements of the S-R ensembles were used: a maximum or direct relation between the stimuli and the responses (usually considered to involve the greatest "compatibility") a mirrored or reversed relationship, and a random assignment of responses to stimuli. With either the maximum direct or mirrored conditions, the spatial two-dimensional arrangement produced lower response times than all other groups. Performance was generally poor with the random condition but here there was significantly better performance with the two symbolic coding sets than with either of the spatial sets. Differences appeared to be relatively permanent over time.

Fitts and Seeger (1953), the first investigators to use the term "S-R Compatibility", made the following comments on the general notion of S-R compatibility:

It is not permissible to conclude that any particular set of stimuli, or set of responses, will provide a high rate of information transfer; it is the ensemble of S-R combinations that must be considered.

It appears that it is very difficult for Ss to learn to deal effectively with the information (uncertainties) characteristic of a specific situation, if these uncertainties are different from the more general set of probabilities which have been learned in similar life situations.

Thus these investigators felt that the stability of these differences reflects the fact that learning of S-R conditions which violate expectations is subjected to interference from competing habits which have been reinforced over a period of many years.

Garvey and Knowles (1954) had subjects push buttons in a matrix in response to the turning on of a light in a matrix. They examined response times associated with various display-control relationships. The six displaycontrol configurations they used were: 1) a 10 x 10 matrix with a button beside each light; 2) a 10 x 2 matrix with a button beside each light; 3) a 10 x 10 matrix of lights above a 10 x 10 matrix of buttons; 4) a 10 x 2 matrix of lights with a 10 x 2 matrix of buttons: 5) a 10 x 10 matrix of lights with a 10 x 2 matrix of buttons; 6) a 10 x 2 matrix of lights with a 10 x 10 matrix of buttons. They found that mean response times differed significantly for all systems. The systems. listed best to worst (i.e., in decreasing order of compatibility), were: 1, 2, 3, 4, 6, and 5. The investigators also found no change in relative efficiency over practice. When they added a secondary task (counting clicks), they found that there was no significant effect on the per-

formance of the primary task using conditions 1 and 2, but the other conditions were affected, with the least efficient systems showing the greatest effect. The following year, Garvey and Mitnick (1955) took the last four display-control configurations listed above and added lines breaking the matrices into segments (in the previous study the lights and buttons were mounted on plain backgrounds). They again found the best performance "when display and control --- matched so that an isomorphic relationship existed between stimulus and response sets; i.e., the spatial arrangement of the stimulus elements on the display was identical to that of the response elements on the control." They found that where the internal interference among the stimuli is the greatest, additional spatial references (the lines added to the matrices) can enhance the efficiency of the display-control system, but they also found that the addition of an "excessive" number of spatial references may degrade performance.

Operator Load as a Function of S-R Compatibility.

The studies described above have examined S-R compatibility in terms of its effect on response times. Various degrees of spatial and symbolic relationship between displays and controls have shown response times to decrease as the S-R relations in question approach relevant population sterotypes (greatest compatibility). These studies, however, have generally followed the format of reaction

time experiments, in which the subject is set for each new stimulus, and there is no strict pacing imposed on the sequence of stimuli.

The Present Study

The present study hypothesized that in self-paced sequential tasks involving different degrees of S-R compatibility, S-R configurations resulting in longer response times (and inferred to have less compatibility) would produce a slower pacing rate. If the pacing were forced, and if a constant presentation rate were used over all conditions which would allow even the most difficult configuration to be handled adequately, then it should be demonstrable that varying only the degree of S-R compatibility produces a concomitant variation in what we have earlier called "operator load". That is, given a constant pacing rate over all conditions, the task condition whose individual presentations produce the longer response times will leave the operator with less "free" time in which to carry on an additional task, if the additional task is such that it must be timeshared with the primary task.

It appears then that the degree of operator loading of a task might be controlled by varying the task along the dimension of S-R compatibility. The basic function of the task and the amount of information transmitted may remain constant, and only the relationship between the display and the controls would vary.

A secondary task which presented information not readily combined or coded with the primary task information (and therefore requiring time-sharing) could then be used as an index of the operator loading if priorities were assigned to the tasks in such a way that the subjects tried to perform the primary task without error and devoted only whatever time was left over to the performance (imperfect, if necessary) of the secondary task.

auditory inspection task as a means of examining possible differences in S-R compatibility of four different S-R configurations in a paced light matrix-pushbutton primary task. It was hypothesized that the S-R configurations represented different degrees of S-R compatibility, which would be reflected, through changes in operator loading, as different error rates on the simultaneously performed auditory task. It would be expected that secondary task error rates would be an inverse function of compatibility in the primary task.

SECTION II

METHOD

Subjects

The forty subjects who participated in the study were volunteers from introductory psychology classes. The subjects ranged in age from 17 to 20; mean age was 18.6 years. There were 20 males and 20 females; five males and five females were randomly assigned to each of the four experimental conditions.

Design

For the four experimental conditions four different S-R configurations, representing different degrees of S-R compatibility, were imposed on a visual-motor task. The relation (described in detail below) between patterns of extinguished lights in a matrix and the position of pushbuttons used to relight the lights was varied. The simultaneously performed secondary task (also described in detail below), whose error rate was to be used as an index of difficulty of the primary task, was an auditory inspection task in which the subject heard (through head-phones) a tape-recorded series of numbers and made verbal responses to new digits appearing in the series.

A single-classification analysis of variance design was used, with four treatment groups (visual task S-R configurations) of ten subjects each. Each subject was given four trials of 100 auditory and simultaneously,

240 visual presentations each, and errors were scored for both the auditory and visual tasks. An overall score for each of the tasks, mean errors per trial, was then computed for each subject, and these scores were used in the analyses of variance. Separate analyses of variance were done for the two tasks. Although the auditory task scores were intended to be the index of S-R compatibility, an analysis was also performed on the visual task scores to see if there were any differences in the extent (generally small) to which the visual task performance was degraded by the addition of the second task. Following these analyses of variance using the overall means, differences within each trial were examined by means of additional analyses of variance and "critical difference" tests, and group mean error scores for each trial were plotted for the four trials to provide a graphic representation of practice effects.

The Tasks

Primary Task. This task, whose S-R compatibility (and associated operator load) was varied, was a visual-motor task in which four different stimuli were presented in a programmed random sequence on a 3 x 4 matrix of red panel lights. Each stimulus was the location or pattern of one or more lights which were extinguished from the otherwise fully lighted matrix. Any particular light in the matrix could be a member of only one pattern. The

response to any one of these four patterns was made by pushing one of four pushbuttons, arranged horizontally on a small box in front of the subject in such a way that the subject could make all responses with his right hand, one finger per button. A correct response relighted the extinguished stimulus lights, producing a fully lighted matrix, which remained lighted until the next stimulus in the random sequence occurred (stimuli were presented at a rate of approximately one every 1.7 seconds). An incorrect response or the simultaneous depression of two or more buttons failed to relight the stimulus lights. All responses were recorded automatically on paper tape by a six-channel event recorder. The arrangement of the pushbuttons (i.e., horizontally spaced from left to right in front of S so that each finger of the right hand manipulated a single button) remained the same for all S-R conditions. The four sets of S-R relations are illustrated in Figure 1. The conditions are as follows:

1. Symbolic. In this condition, the actual number of extinguished lights specified the response (see Figure 1). That is, the first or left button (pushed by the index finger) was pushed in response to a single extinguished light (always the third light from the left in the bottom row); the second button (middle finger) was the correct response to the two extinguished lights pictured in Figure 1, etc.

		Button Number					(L	(Left-Right)								
Stimulus		1			2				3				4			
Pattern:	0	0	0	0	0	0	0	0	0	0	0	0	•	•	•	0
Symbolic	0	0	0	0	0	0	0	0	•	•	0	•	0	0	•	0
Symbolic .	0	0	•	0	٥	•	0	•	0	0	0	0	0	0	0	0
	•	0	0	O	0	•	0	0	0	0	•	0	0	0	0	•
Spatial, Direct	•	0	0	0	٥	•	0	0	0	0	•	0	0	0	0	•
	•	0	0	0	0	•	0	0	0	0	•	0	0	0	0	•
Spatial, Reverse	0	0	٥	•	0	0	•	0	0	•	0	0	•	0	0	0
spatial, heverse	0	0	0	•	0	0	•	0	0	•	0	0	•	0	0	0
	0	0	0	•	٥	0	•	0	0	•	0	0	•	0	0	0
Random	0	0	•	0	•	0	0	0	0	0	0	•	0	•	0	0
	•	•	0	0	0	0	•	0	0	0	0	•	٥	٥	0	0
	0	0	0	0	0	•	0	0	0	0	•	0	•	0	0	•
	•		_	ted ngu:	ish	ed										

Figure 1. Stimulus Configurations Corresponding to Response Button Positions for Four Stimulus Conditions.

tatnad<u>l</u> Heriado harado

. .

ar Intita

5110

troni" ,inime

e abawa 1 .iaitaa.

is complementable transfer to the Complete Compl

and of Electric dill states.

- vancityProl a i rut

- 2. Spatial, Direct. Here, the horizontal location of columns of three lights each was directly related to the horizontal positions of the buttons. The left button controlled the left column, the right button controlled the right column, etc.
- 3. Spatial, Reverse. Here, the relation in the second condition was reversed; the left button controlled the right column, the right button controlled the left column, etc.
- 4. Random. In this condition, the four patterns of three lights each pictured in Figure 1 were randomly arranged, and were arbitrarily assigned to the buttons in the way indicated, with no obvious spatial or symbolic relation to the response buttons.

The sequence of presentation of the four stimuli within a particular condition was determined by a random number table, which was used to set up a random sequence of 200 steps, in a closed loop which automatically repeated without a break. When the primary and secondary tasks were performed together, the primary task was automatically paced at a rate of approximately one presentation every 1.7 seconds. This rate was found in a preliminary investigation to allow performance with no errors (after practice) with any of the four S-R configurations, when the task was performed alone.

The lights were arranged in a matrix with four columns of three lights each. The lights were mounted on a board, 8" x 8", and the board was fitted with a hood or shade 7" deep to keep the overhead lights in the room from shining directly on the display. The whole display was mounted at eye level, six feet directly in front of the seated subject, who looked straight ahead at the display and responded to the patterns of extinguished lights by pushing the appropriate buttons on the table in front of him.

Secondary Task. The secondary task was an auditory inspection task. The subject heard, through headphones, a tape-recorded series of four-digit numbers, each number differing from the preceding one in only one digit. For example, the new digit in each of the numbers below is underlined:

The new digit is:

1 2 3 4	
1 2 5 4	5
1 3 5 4	3
1 3 5 <u>8</u>	8
7 3 5 8	7

As the subject listened to the numbers, his correct response, in each case, would be to call out the new digit (the one underlined in each number above) as soon as he had heard the whole four-digit number. The numbers were read at the rate of two digits per second, and

there was a two second pause between numbers, during which time the subject was instructed to say aloud his response. Thus the subject heard and responded to a new four-digit number every four seconds. This pacing rate was the same as that used in an auditory inspection task described by Poulton (1960), but the present study used four-digit numbers, while Poulton used eight-digit numbers. A pilot study was conducted to determine a workable number length for the present study, and it was found that eight- and six-digit numbers yielded a rather high error rate even after considerable practice with the auditory task alone. The four-digit numbers, however, were quite easy for the subjects to follow, and they were able to perform with an average of fewer than five per cent errors (on the auditory task alone) after relatively little practice.

The numbers used in the auditory inspection task were intended to appear random, but they were produced under many of the same constraints as those used by Poulton (1960) in generating his eight-digit numbers:

- 1. Only the numerals 1 through 8 were used.
- 2. In a set of 32 consecutive four-digit numbers, each of these 8 numerals appeared as a change once in each of the four serial positions.
- 3. Two consecutive new digits could not be the same numeral, nor could the new digit appear in the same

serial position. In addition, the four appearances of each digit in a set of 32 numbers were spaced out so that all four could not appear closely together. Likewise, the eight changes in each of the four serial positions were spaced out so that the new digit did not appear in the same serial position several times in a short interval of time.

- 4. No digit could appear more than twice in any four-digit number; if a digit appeared twice, the serial positions could not be adjacent (e.g., 22 or 88 could not appear anywhere in the number).
- 5. The number at the end of the second set of 32 numbers was identical with the initial number which started the first set of 32. Thus a total of 64 numbers formed a closed loop, which was recorded so that ten identical copies of the loop followed each other without a break.

The experimenter sat to the left of the subject and used a master scoring sheet to tally the subject's correct responses to the auditory task. Each subject was instructed to say his responses aloud, and to speak clearly. He was told that any response that came after the next number began would be scored as incorrect.

Procedure

The subject was given a set of written instructions which he read to himself (these instructions are reproduced in Appendix I). He was directed by the instructions to

stop reading at particular points. Each time he stopped he was allowed to ask questions, and then he was allowed to practice the particular part of the task about which he had just read.

- 1. First, the auditory task was described in detail in the instructions. Then the subject put on the headphones and was given 64 practice presentations on the auditory task. Subjects often made errors on the first few responses, but they generally started performing consistently well after about the first ten. The last fifty of these practice responses were recorded, to make sure that the subject was not having any unusual trouble. All subjects were given the same number of practice trials.
- 2. Next the visual task was described in the instructions. Then the subject was allowed to practice the visual task at his own rate (i.e., the stimulus did not change until the subject made a correct response), but as he began to learn the responses and to perform accurately, he was encouraged to speed up and perform as rapidly as possible without making errors. Each subject made at least 75 responses: after the first 50 responses (regardless of errors), each subject continued until he made 25 consecutive responses without an error. At the end of this self-paced practice all subjects were performing at a rate well within the automatic presentation rate used in the rest of the study.

- 3. Next the equipment was set to present the stimuli at a constant rate, approximately one stimulus every 1.7 seconds. Every subject then practiced the task with the automatic pacing, for at least 45 responses: after the first 20 responses (regardless of errors), each subject continued until he made 25 consecutive responses without making an error. If the subject made an incorrect response but was able to correct it before the next stimulus was presented, this was scored as one correct response. If a presentation went by with either no response, or an incorrect response only, this was scored as an error.
- 4. Finally, the subject practiced both tasks simultaneously for 35 auditory (83 visual) presentations.
- 5. After the subject had practiced the combined tasks he was told that from that point on all responses would be scored and that he would be allowed to take occasional breaks to prevent his becoming too tired. He was then given four trials, on the combined tasks, of 100 auditory and 240 visual presentations each, with breaks of about five minutes between trials.
- 6. At the end, each subject was scored on an additional 50 presentations of the auditory task alone.

 Scoring of Responses

For the auditory task, number of errors (late response, no response, or incorrect response) per trial (100 presentations per trial) were scored by the experimenter,

using a master list to check the verbal responses. Performance on the auditory task alone, both before and after the combined task, was also scored. Errors were scored for the last 50 of the 64 practice presentations, and also for the final 50 presentations of the auditory task alone at the end of the experiment.

For the visual task, number of errors (allowing a stimulus to pass without making a correct response) per trial (240 presentations per trial) were scored, using the paper tape record of the visual task performance.

Informal comments by the subjects were also recorded, and some, which proved to be of interest, are factored into the discussion of results, below.

SECTION III

RESULTS

Auditory Task

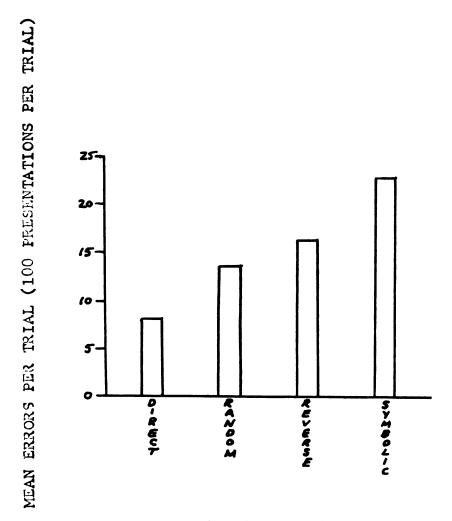
Mean Errors Per Trial. A score of mean errors per trial was calculated for each subject. A single-classification analysis of variance yielded the following results:

Source	df	<u>ss</u>	MS	<u>F</u>	P
S-R Configuration	3	1127.01	375.67	5.01	<.01
Within Subjects	<u>36</u>	2694.94	74.86		
Total	39	3821.95			

The overall means for the subject groups are shown graphically in Figure 2, and the mean differences are shown in tabular form in Table 1. Table 1 also includes a "Critical Difference" (Lindquist, 1953) which was computed after the F in the analysis of variance was found to be significant at the .01 level. Note that the only mean difference which was significant at the .01 level was the difference between the direct and symbolic groups.

Practice Effects. The auditory task was performed alone, both at the beginning and at the end of the experiment. The overall mean error rate for these two auditoryalone trials (50 presentations each) was 4.5 per cent before the combined tasks, and 2.25 per cent after, for a net reduction in errors of 2.25 per cent from the beginning to the end of the experiment.

Within each subject group the combined-task error scores were broken out for each trial, and the changes



S-R COMPATIBILITY CONDITION

Figure 2. Effect of S-R Compatibility on Mean Auditory
Task Errors Per Trial.

TABLE 1

GROUP MEAN DIFFERENCES AND CRITICAL DIFFERENCE
FOR SUBJECT MEAN AUDITORY ERRORS PER TRIAL

	Random	Reverse	Symbolic
Direct	5.675	8.275	14.775**
Random		2.600	9.100
Reverse			6.500

Critical Difference (p < .01) = 10.5 ** p < .01

in mean error rate over trials are plotted in Figure 3.

Separate analyses of variance were performed on these data for each trial. The results of the analyses were as follows:

Trial 1:

Source	df	SS	MS	<u>F</u>	P
S-R Configuration	3	2108.90	702.97	4.47	<.01
Within Subjects	<u>36</u>	5663.90	157.33		
Total	39	7772.80			
Trial 2:					
S-R Configuration	3	931.88	310.63	4.01	<.05
Within Subjects	<u>36</u>	2790.50	77.51		
Total	39	3722.38			
Trial 3:					
S-R Configuration	3	665.00	221.00	3.27	<.05
Within Subjects	<u>36</u>	2432.90	67.58		
Total	39	3097.90			
Trial 4:					
S-R Configuration	3	1190.90	396.97	5.18	<.01
Within Subjects	<u>36</u>	2757,00	76.58		
Total	39	3947.90			

Critical differences were computed at the level of significance of each F. In each of the four trials, the only significant mean difference was the difference between the direct condition and the symbolic condition.

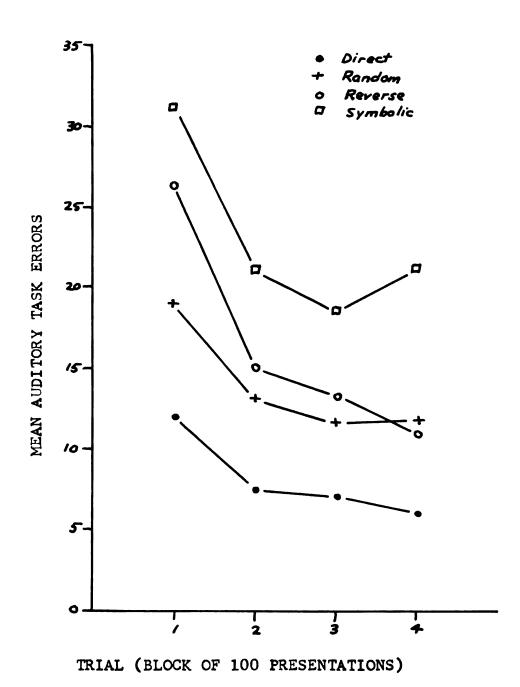


Figure 3. Effect of Trials on Mean Auditory
Task Errors.

Within each subject group a t test for correlated observations (Winer, 1962) was applied to the mean error rates for the first and last trials, to see if the mean error rate for the group changed significantly between the beginning and the end of the experiment. The results of the t tests were as follows:

	\overline{X} , Trial 1	\overline{X} , Trial 4	Difference	<u>t</u>	P
Direct	12.0	6.0	6.0	4.05	<.01
Reverse	26.3	11.0	15,3	5.26	<.01
Random	18.9	11.7	7.2	2.30	<. 05
Symbolic	31.1	21.1	10.0	2.77	<. 05

It may be seen that each subject group showed an improvement in performance over the trials significant at or beyond the .05 level.

Visual Task

Mean Errors Per Trial. A score of mean visual task errors per trial was calculated for each subject. A single-classification analysis of variance yielded the following results:

Source	<u>df</u>	<u>ss</u>	MS	<u>F</u>	P
S-R Configuration	3	189.97	63.32	4.39	<.01
Within Subjects	<u>36</u>	518,90	14.41		
Total	39	708.87			

The overall means for the subject groups are shown graphically in Figure 4, and the mean differences are shown in Table 2. Table 2 also includes a critical difference which was computed after the F in the analysis of variance above

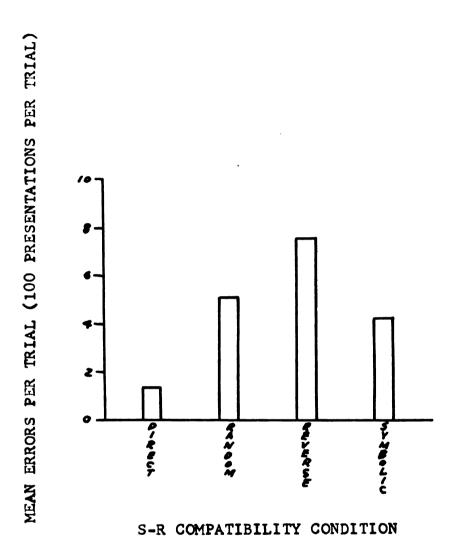


Figure 4. Effect of S-R Compatibility on Mean
Visual Task Errors Per Trial.

sorm order million, seed that

Second the Mills with semmed them to the first the second of the second

TABLE 2

GROUP MEAN DIFFERENCES AND CRITICAL DIFFERENCE
FOR SUBJECT MEAN VISUAL ERRORS PER TRIAL

	Random	Reverse	Symbolic
Direct	3.65	6.10**	2.80
Random		2.45	0.85
Reverse			3.30

Critical Difference (.01) = 4.61

** p <.01

was found to be significant at the .01 level. Note that the only mean difference which was significant at the .01 level was the difference between the direct and the reverse groups.

Practice Effects. Within each subject group, the combined task error scores were broken out for each trial, and the changes in mean error rate over trials are plotted in Figure 5. Separate analyses of variance were then performed for each trial. The results of the analyses were as follows:

Trial 1:

Source	<u>df</u>	<u>ss</u>	MS	<u>F</u>	P
S-R Configuration	3	564.20	188.07	4.93	<.01
Within Subjects	<u>36</u>	1371,80	38.11		
Total	39	1936.00			
Trial 2:					
S-R Configuration	3	119.40	39.80	1.82	n.s.
Within Subjects	<u>36</u>	785.00	21.81		
Total	39	904.40			
Trial 3:					
S-R Configuration	3	102.60	34.20	3.16	<. 05
Within Subjects	<u>36</u>	389,80	10.83		
Total	39	492.40			
Trial 4:					
S-R Configuration	3	122.08	40.69	2.11	n.s.
Within Subjects	<u>36</u>	692.70	19.24		
Total	39	814.78			

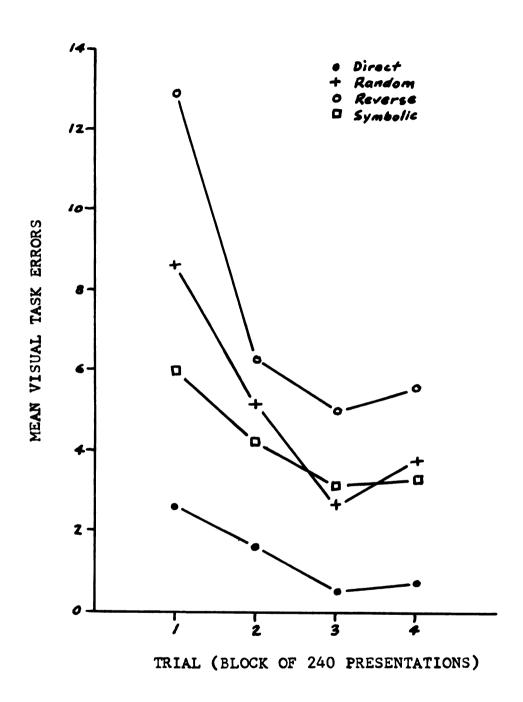


Figure 5. Effect of Trials on Mean Visual Task Errors

through the truth of the control of

Critical differences were computed at the level of significance of each F. Note that there were no significant differences among the subject groups for trials 2 and 4. For trials 1 and 3, the only significant mean difference was the difference between the direct condition and the reverse condition.

Within each subject group, the t test for correlated observations was applied to the mean error rates for the first and last trials, to see if there was a significant change in mean error rate between the beginning and the end of the experiment. The results of the t tests were as follows:

	\overline{X} , Trial 1	\overline{X} , Trial 4	Difference	t	P
Direct	2.6	0.7	1.9	1.62	n.s.
Reverse	12.9	5.6	7.3	2.54	< .05
Random	8.5	3.7	4.8	2.23	n.s.
Symbolic	6.0	3.3	2.7	1.80	n.s.

It may be seen that only the Reverse group showed a significant improvement in performance over the trials (p < .05).

SECTION IV

DISCUSSION AND SUMMARY

Auditory Task

The results indicate that the secondary auditory inspection task, used as an index of operator loading, was sensitive to changes in the stimulus-response configurations of the primary task.

Using the index of operator loading also as an index of the S-R compatibility of the primary task configurations, the plotted group means in Figure 2 show how the configurations ranked in terms of stimulus-response compatibility. The direct spatial condition, which represented the most "natural" relationship, produced the lowest secondary task error rate. It is therefore concluded that this stimulusresponse combination has the greatest compatibility. The random and reverse conditions came next; these two conditions resulted in about the same auditory error rate, therefore they are judged about equal in compatibility. The symbolic condition resulted in the highest auditory error rate, and therefore (by the present definition) the least compatibility. The reader will recall that the only significant mean difference was between the direct and symbolic conditions.

The relative position of the symbolic condition may, however, be an artifact of the situation. Subjects reported experiencing interference between the numbers used as mediators in the symbolic visual task and the numerical

verbal responses which they made to the auditory task. Eight of the ten subjects in the group reported that they occasionally found themselves saying aloud the number that occurred in the visual task at the instant they began to make the verbal response to the auditory task. Instead of making the appropriate verbal response, they would say the number that corresponded to the next visual response. No subject, upon questioning, reported any interference in the other direction, i.e., the auditory responses never intruded into the visual task. The two subjects who experienced no interference reported that they disregarded the numerical mediator for the visual patterns, and responded directly to the patterns. When these subjects' data were analyzed, it was found that they had the two lowest mean error scores.

This suggests that the secondary task selected for use in the present study violated the concept of intertask compatibility, mentioned earlier as being one of the desirable characteristics of a secondary task used as a measuring device. If there had been no such interaction between the primary and secondary tasks, the relative position of the symbolic condition might have been different (e.g., if the lights had "spelled out" A,B,C,D, or if a non-numerical secondary task had been used).

The analyses of the data within each trial (see Figure 3) showed that throughout the course of the experi-

ment the difference between the direct and symbolic groups remained the only significant difference. Each of the four groups showed an improvement in performance over the trials, but the greatest improvement was between the first and second trials.

It is possible that the curves would stabilize if more trials were added, but it is also possible that fatigue might begin to affect performance. The performance of the combined tasks was very demanding, even for the "easiest" S-R condition, and nearly all subjects volunteered comments to this effect. Although each combined task trial lasted only 6.7 minutes, subjects were typically tired and quite ready for the breaks between trials. Visual Task

Although an attempt was made to select a secondary task which would not interfere with the performance of the primary task, some disruption (albeit in minute amounts) of the primary task performance nevertheless occurred.

The means plotted in Figure 4 show that the rank order of the performance degradation was not the same as that for the auditory task. The direct condition was again the "best" (least affected by the addition of the secondary task), but here the reverse condition was affected the most, and the symbolic condition turned up second best.

Comments by the subjects support the finding that the reverse condition was the hardest for performing the visual

task. The subjects all indicated that they understood the S-R relationship perfectly, but many said that they were seriously handicapped by the tendency to make the "direct" response, i.e., to push the button whose horizontal position was the same as the horizontal position of the extinguished light column. (One subject, during the practice session, asked if he could rotate the box on which the buttons were mounted so that the button positions would correspond to the light positions!)

As for the symbolic condition, its appearance as second best is backed by the subjects' earlier comments that the visual symbolic task interfered with the auditory task, but not vice-versa. One might speculate, on the basis of the low error scores for the symbolic visual task, that if the secondary task had not interacted with the symbolic visual condition, the ranking of the S-R conditions as indicated by the secondary task scores might more closely resemble the ranking observed here for the visual task.

One might expect that, as the difficulty of the combined tasks increased (and if the <u>nature</u> of the difficulty remained constant), the ranking of the difficulty-causing conditions, as reflected in performance, would be the same as reflected in <u>either</u> task score (i.e., performance of both tasks would suffer, and the suffering would be highly correlated).

Since the visual task appears not to have been affected by the addition of the secondary task, the rank order of difficulty of the visual task may be the more valid of the two measures, and would be more likely to remain unchanged if this visual task were combined with some other non-interfering secondary task.

The differences among the subject groups were only significant for trials 1 and 3, and within these trials, only the differences between the direct and the reverse conditions were significant at or beyond the .05 level.

Only the reverse condition showed a significant change in performance over the trials, indicating that the subjects had just about reached asymptotic performance by the beginning of the first trial.

It should be noted that the observed mean visual errors represent a very small proportion of the total responses made in a given trial. The entire range of mean errors, expressed as a per cent of total presentations per trial, was 0.56 per cent to 3.10 per cent: very low when compared to the auditory task range of 8.13 per cent to 22.90 per cent.

Summary

Although the secondary auditory task was sensitive to changes in primary task S-R configurations, an interaction was revealed between the numerical mediator used in the symbolic visual task and the numerical verbal responses to the auditory task. This interaction took

the form of visual (numerical) responses intruding into the auditory task responses, but not vice-versa. The validity of the rank order of S-R compatibility as defined by this particular secondary task error rate was therefore questioned, because it was felt that the nature of its interference with the primary task disqualified this secondary task as a good tool for measuring workload with this particular primary task.

Although there were relatively few errors on the visual task, it was determined that there were statistically significant differences in the extent to which the visual task was disrupted by the addition of the secondary task. The direct condition was the best (as in the auditory task). The reverse condition was the most difficult, and the subjects reported that it was so because it violated their response preference or expectation.

APPENDIX I

INSTRUCTIONS TO SUBJECTS

Read Carefully ...

During this experiment you will learn to perform two different tasks: 1) An auditory task, in which you will listen to a tape recording and give your responses verbally, and 2) A visual task, involving the red lights and pushbuttons in front of you.

First you will practice the auditory task alone; then you will practice the visual task alone; finally, you will perform both tasks at the same time. You will be given short rests periodically to keep you from becoming too tired.

Auditory Task

This task is called an auditory inspection task, because you will be inspecting (by listening) a series of numbers, and reporting how each number differs from the previous one.

You will hear a taped series of numbers. Each number differs from the one before it by only one digit.

The numbers you will hear will all be four-digit numbers. For example:

if you hear:

your response will be:

1	2	3	4
1	2	<u>5</u>	
1	3	5	4
	_		3

1 3 5 8 7 3 5 8

The new digit in each number is underlined. This is what your response will be. You will just hear a long series of numbers like the above, with each number having one digit different from the previous number. Your job will be to call out the new digit in each number. There will be a short pause after the reading of each number. This is the time when you should make your response. If you make the response before the number is through being read, you risk forgetting the last few digits. If your response is late--after the next number starts--it will be scored as an error whether you say the right digit or not. So try to wait and say the response during the pause, even if the new digit occurs early in the number.

Speak your responses clearly, so the experimenter can hear you. Don't mumble if you are not sure of the correct response. Speak up even if you feel it's a guess, since a numble may be scored as an error.

DO YOU HAVE ANY QUESTIONS?

(Stop reading each time you come to a line of stars.)

Visual Task

In the light display, there are four different groups of lights which may go off. Only one of these groups will go off at a time. Your task is to relight the group of extinguished lights by pushing the appropriate button.

Which button you push depends on which group of lights goes off. The card over the buttons shows the pattern of extinguished lights that each button controls.

When you push the correct button, the lights all come on. When you push a wrong button, nothing happens, and this is scored as an error. The four light patterns will be presented in a random sequence -- this means that sometimes the same pattern may pop up two or three times in a row.

You will now have a practice session to familiarize yourself with the patterns and their corresponding responses. You may set your own pace, but try to go as quickly as you can without making errors.

DO YOU HAVE ANY QUESTIONS?

Visual Task (continued)

Now that you are familiar with the correct responses, the machine will be set to present the patterns at a constant rate. If you fail to respond within the given time, the machine will automatically go on to the next presentation, and this will be counted as an error. However, the rate is slow enough that you should be able to perform the task easily.

You will now have a practice session. The automatic pacing will be the only difference between this session and the last one. Again, concentrate on making as few errors as possible.

DO YOU HAVE ANY QUESTIONS? ********************* Combined Tasks

You will now have a practice session during which you will perform both tasks at the same time. This is, of course, more difficult than doing either task by itself, and you may make some errors. From now on the most important task is the visual task. If you find yourself unable to perform both tasks without errors, concentrate on maintaining the highest accuracy on the visual task. In other words, try to perform the auditory task as well as possible WITHOUT sacrificing performance on the visual task.

REFERENCES

- Bahrick, H. P., Noble, M., and Fitts, P. M. Extra-task performance as a measure of learning a primary task. J. exp. Psychol., 1954, 48, 296-302.
- Bahrick, H. P., and Shelley, C. Time sharing as an index of automatization. <u>J. exp. Psychol.</u>, 1958, 56, 288-293.
- Benson, A. J., Huddleston, J. H. F., and Rolfe, J. M. A psychophysiological study of compensatory tracking on a digital display. <u>Human Factors</u>, 1965, 7, 457-472.
- Brown, I. D. Measuring the spare mental capacity of car drivers by a subsidiary auditory task. Ergonomics, 1962, 5, 247-250.
- Brown, I. D. and Poulton, E. C. Measuring the spare mental capacity of car drivers by a subsidiary task. Ergonomics, 1961, 4, 35-40.
- Fitts, P. M., and Deininger, R. L. S-R compatibility: Correspondence among paired elements within stimulus and response codes. J. exp. Psychol, 1954, 48, 483-492.
- Fitts, P. M., and Seeger, C. M. S-R compatibility: Spatial characteristics of stimulus and response Codes. J. exp. Psychol., 1953, 46, 199-210.
- Garvey, W. D., and Knowles, W. B. Response time patterns associated with various display-control relationships. J. exp. Psychol., 1954, 47, 315-322.
- Garvey, W. D., and Mitnick, L. L. Effect of additional spatial references on display-control efficiency. J. exp. Psychol., 1955, 50, 276-282.
- Garvey, W. D., and Taylor, F. V. Interactions among operator variables, system dynamics, and task-induced stress. J. appl. Psychol., 1959, 43, 79-85.
- Knowles, W. B. Operator loading tasks. <u>Human Factors</u>, 1963, 5, 155-161.
- Lindquist, E. F. Design and analysis of experiments in psychology and education. Houghton Mifflin Co., Boston: 1953, 90-96
- McCormick, E. J. Human Factors Engineering. McGraw-Hill, New York: 1964.

- Poulton, E. C. The optimal perceptual load in a paced auditory inspection task. <u>Brit. J. Psychol.</u>, 1960, 51, 127-139.
- Ross, S., Shepp, B. E., and Andrews, T. G. Response preferences in display-control relationships.

 <u>J. appl. Psychol.</u>, 1955, 38, 425-428.
- Winer, B. J. Statistical principles in experimental design. McGraw-Hill, New York: 1962, 39-41.

MICHIGAN STATE UNIVERSITY LIBRARIES

3 1293 03142 3910