SUBJECTIVE SCALING OF MENTAL WORKLOAD IN A MULTI-TASK ENVIRONMENT

by

BAHMAN DARYANIAN

B.S., Massachusetts Institute of Technology
(1977)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

(February 1980)

Signature of Author . . . Signature redacted . . . . .
Department of Mechanical Engineering,
February , 1980

Certified by . . . Signature redacted . . . .
Thesis Supervisor

Signature redacted

Accepted by . . . .
Chairman, Department Committee on Graduate Students

SUBJECTIVE SCALING OF MENTAL WORKLOAD IN A MULTI-TASK ENVIRONMENT

by

BAHMAN DARYANIAN

Submitted to the Department of Mechanical Engineering on February , 1980 in partial fulfillment of the requirements for the Degree of Master of Science.

ABSTRACT

Interest in defining and measuring mental workload is a recent development, the purpose of which is to predict how mental involvement of aircraft pilots and other human operators in man-machine interfaces affects performance. This work examines mental workload in the context of a simulated multi-task decision environment through construction of subjective scales of mental workload and analyses of subjective data.

The experimental paradigm used for multi-task decision environment was originally developed by Kamil Tulga. Experimental subjects were asked to work on 27 different decision-making cases and judge them with respect to their associated mental workload on paired compaisons basis. The decision-making cases consisted of various combinations of three experimentally controlled variables: interarrival time between tasks, task speed, and operator productivity, where each variable could be set at any of a high, medium, or low level values. Thurstone's Law of comparative judgment was employed in order to construct interval scales of subjective mental workload from paired comparisons data.

Analyses of variance and other statistical analyses were performed on the data. It was found that, in general, a lower number of tasks-to-be-processed per unit time (a condition associated with longer interarrival times) results in a lower mental workload, a higher consistency of judgments within a subject, a higher degree of agreement among the subjects, and larger distances between the cases on the Thurstone scale of subjective workload. It was also

found that the control variables and their interaction had dissimilar effects on the variation of mental workload for different levels of interarrival time.

Thesis Supervisor: Thomas B. Sheridan
Title: Professor of Engineering and Applied Psychology

ACKNOWLEDGEMENTS

I had the privilege of doing research under the supervision of Professor Thomas B. Sheridan at the Man-Machine Systems Laboratory at M.I.T. I thank him for his advice and patience during every unfolding stage of this work.

I also thank Prudence Young for her quality typing of this manuscript.

This project was supported by grant NSG 2118 from the NASA Ames Research Center, Moffett Field, CA, 94035. Thomas E. Wempe was the grant monitor.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

TABLE OF CONTENTS (Cont.)

TABLE OF CONTENTS (Cont.)

List of Figures

List of Figures (Cont.)

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Prelude

The industrial revolution organized men into collective operators of industrial tools of production. The last three centuries has witnessed the progressive transformation of these industrial tools of production into more efficient, and also more complex, industrial systems. The tasks of human operators have, also progressively, changed with these transformations. In the past the primitive dynamic systems required the human operator to perform as a continuous in-the-loop controller. The more recent introduction of automatic and pre-programmed subsystems in many technological areas, now require the human operators - such as pilots (Palmer, 1975), air-traffic controllers (Wasson, 1977), and controllers in the nuclear-reactor and chemical-process industries - to assume the roles of monitor, supervisor, and decision maker in addition to that of a controller.

The increasing complexity of operator's responsibilities has introduced new problems in the design, development, and performance assessment of systems. The best example is the ever increasing refinement and improvement of avionic systems - commercial and military aircraft - with subsequent increase in complexity of subsystems with which human operators and pilots have to deal. A constraining factor in design and development of such systems is the limited capability of human operators who supervise and control them.

Apparently, there exists a need for establishment of some criteria for an acceptable degree of compatibility in interaction between operator and system (man and machine). The earliest candidate as a subject of investigation in this area has been the study of performance of a system. But, it has been reasoned that, although the study of performance of a system may indicate the limits of the abilities of its human operators, it does not reflect the intensity of demand that is placed on operators (Sheridan and Simpson 1979). An unskilled pilot under extreme mental strain may perform as well as a highly skilled pilot who is doing his daily routine job. The concept of mental workload (MWL) was developed to explain such relative state of interaction between operators and systems.

It is hoped that a good understanding of the phenomena of MWL would enable designers of the new systems to avoid placing unreasonable demands on the operators. Origins of the interest in MWL is also traced back to the novel desire of investigators to predict the failure of a system prior to its actual occurance. It has been postulated that the operator-related failure of a system occurs

when the MWL reaches an unacceptable level.
Therefore, a good understanding of MWL would also help in early identification of operating conditions which result in unacceptable levels of MWL, and consequently, in probable breakdown of the systems.

The new enthusiasm for the question of MWL has been well demonstrated in several international symposia which have been held in the past three years (References 1, 2 and 12).

1.2 Conceptual Frameworks for the Understanding of Mental Workload

The interaction between the human operator (HO) and the sysem takes place at the so-called man-machine interface. Figure 1.la depicts an abstraction of the man-machine interaction. At the interface, the HO monitors the behavior of the system via display panels, indicators, and various instruments. Commands to the system are put in through the control panel that may include joy sticks, control switches, keyboards, etc. The human side of this closed loop system could be modeled as a combination of three interactive subsystems or components. These include sensory, cognitive, and motor components (Sheridan and Simpson 1979).

Depending on the requirements of the system human performance could be classified under three basic categories, namely, information processing, decision making, and control (Sheridan and Ferrel 1974). A routine task carried out by the HO may, infact, include all these three categories to some extent. In recent years the emphasis has shifted from the control tasks to the information processing and decision making tasks. This is parallel to the evolution of the role of the HO from a continuous in-the-loop controller to a supervisor/manager. The crossover model of McRuer (1965) and the optimal control model of Kleinman, Baron and Levison (1971) form the backbone of manual control models. There also exist an extensive literature on supervisory control models of HO (Sheridan, 1970; Sheridan and Johannsen, 1976).

From the comparison of the man-machine system model of Fig. 1.1a with the block diagram of a classical control system in Fig 1.1b we can draw the following analogy: The performance criterion and the human operator together perform the duty of the controller and the final control element of the classical control. As in any control system the important aspect of a man-machine system is the performance. The analogy between the two systems could be extended further without loss of relevance. A very important problem that concerns all the control systems is that the

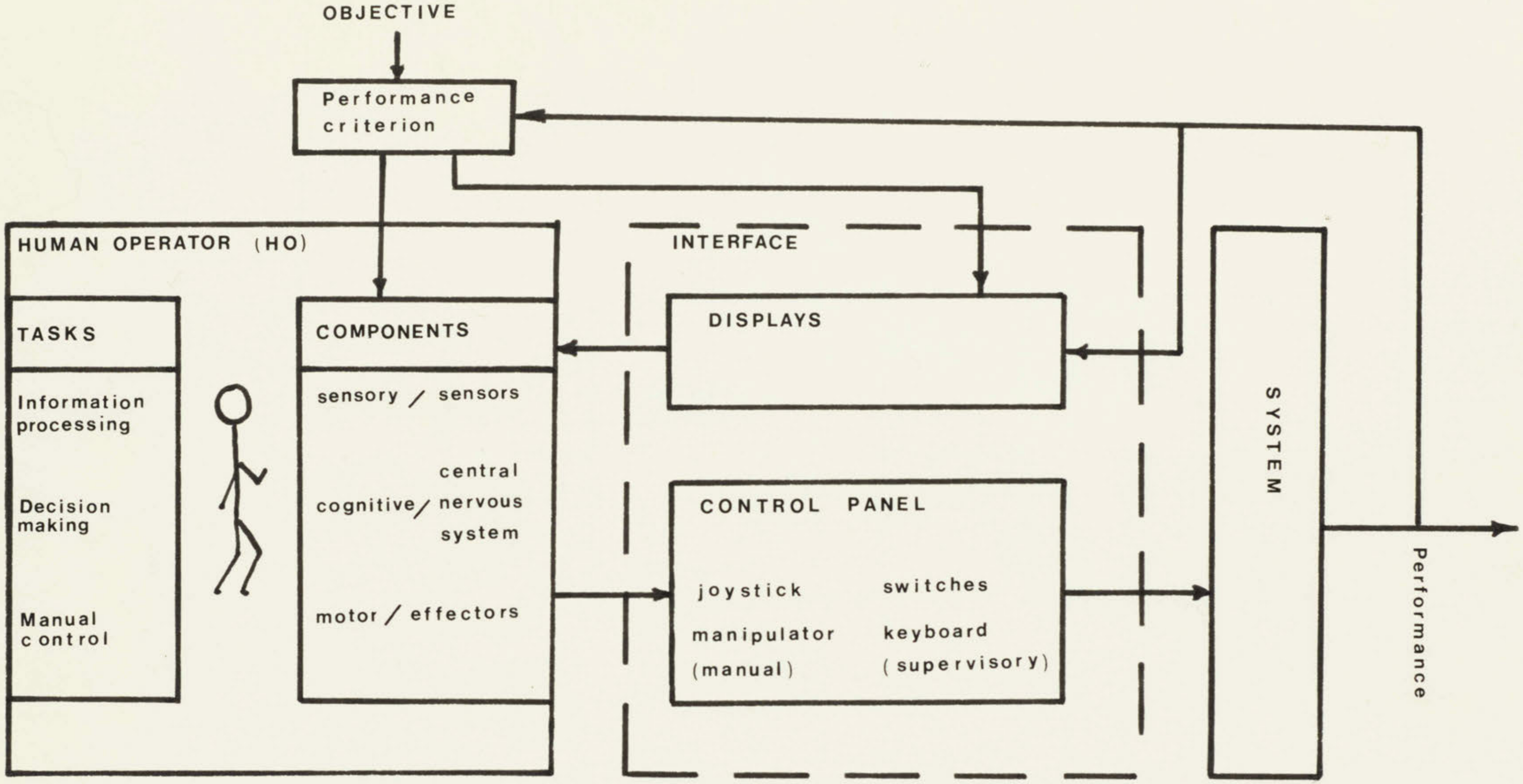
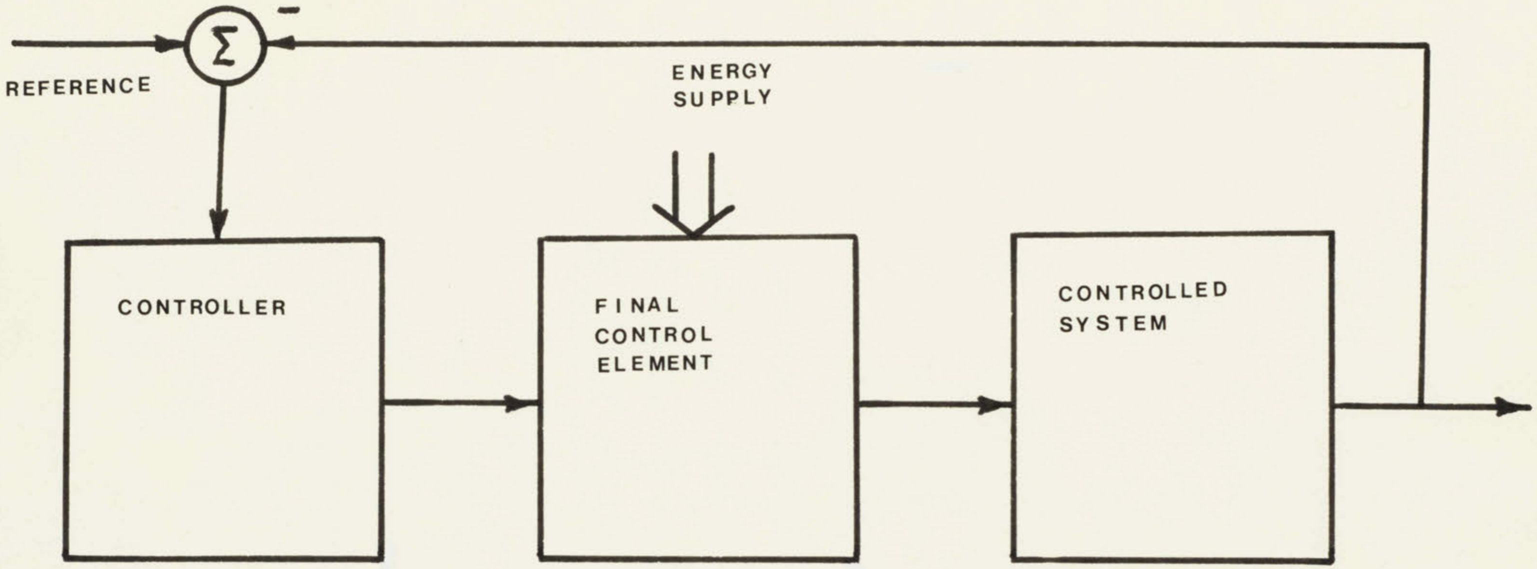


Figure 1.1a Control Diagram for Man-Machine Interaction.



(from reference 14)

Figure 1.1b Control Diagram for an Industrial System performance is ultimately limited by the ability of a final control element to deliver energy to the controlled system (e.g., Phalen, 1977). The concept of the "load" imposed by the controlled system on the final control element can be extended to include a similar situation in man-machine systems. The system to be controlled or supervised, because of its inherent characteristics, imposes some "load" or "demand" on the human operator.

In the light of these developments the concept of mental workload could be used to symbolize the work done in the "head" associated with certain mental activities such as information processing and decision making involving sensors and cognitive components of HO. This is in contrast to physical workload which involves muscles (effectors, motor component) of the HO. Mental workload is closely associated with other mental phenomena such as "thinking", "fatigue", "stress", and the like. However, here the focus is more on the operational (engineering) aspects of such phenomena.

Sheridan and Stassen (1979) suggested six alternative definisions of workload in the context of a control diagram as illustrated in Fig. 1.2. Figure 1.2 also shows the relationship among various measures of workload with respect to the control paradigm.

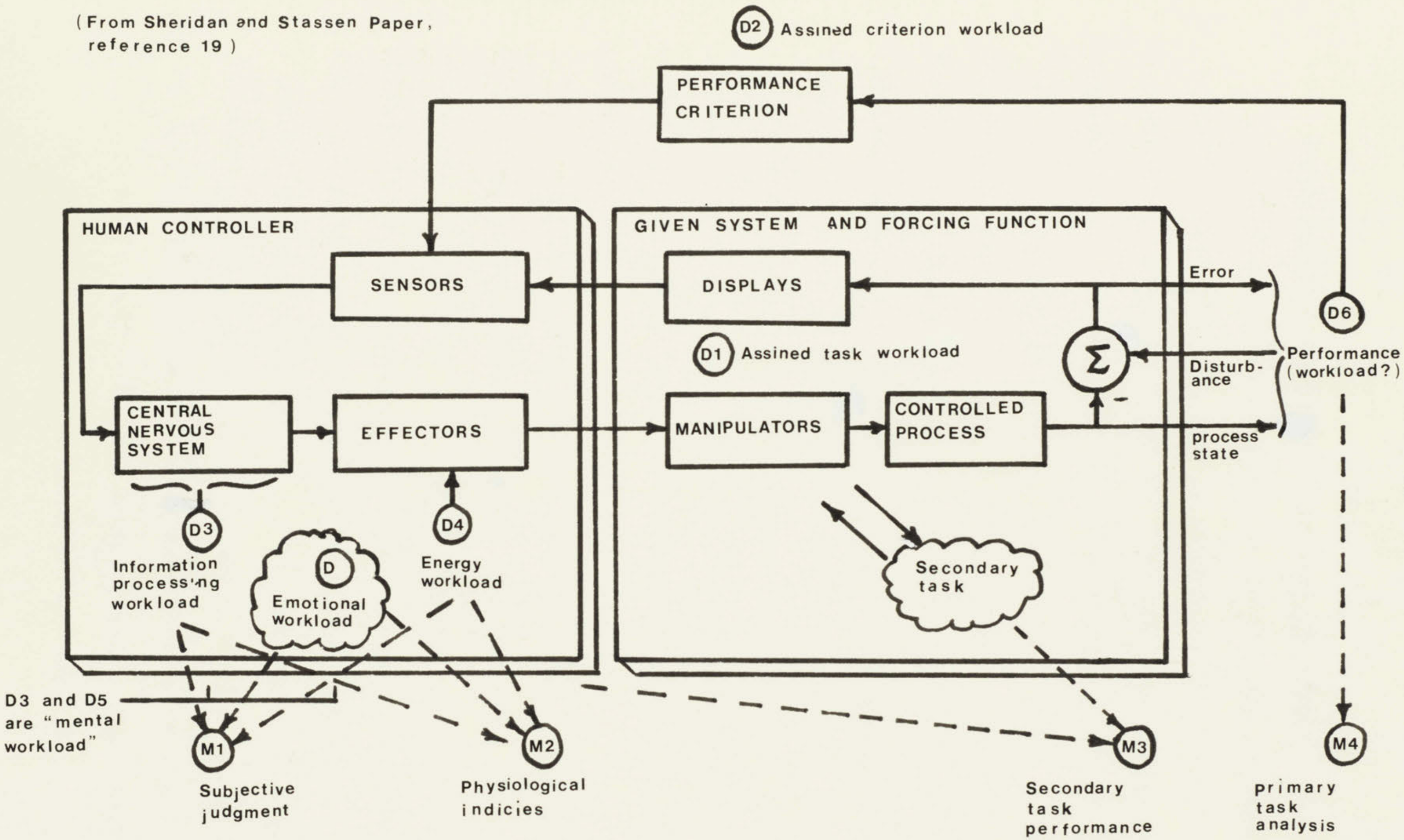


Fig. 1.2 Alternative Definitions of Workload and Performance Illustrated in Control Paradigm.

1.3 Need for an Ideal Experimentally Controlled Environment

The conceptual constructs of the previous section clearly indicate that mental workload is associated with the man-machine interface and as such it should reflect the characteristics of both human operator and the system environment. It may prove to be an impossible task to develop a universal theory of MWL. A prior need will be the comprehensive understanding of both the HO and system characteristics. Even if variations of operator characteristics were to be classified exhaustively, it may be impossible to do the same for all actual supervisory and control environments. The only hope rests on the belief that the findings for specific cases could be generalized to include similar situations.

As in any scientific pursuit the best starting point would be experiments with the simplest supervisory and control environments, where proper manipulation of the variables of the system for investigative purposes is possible. This not only would enable the investigator to design a properly controlled experiment, but also will decrease the complexity of analysis. I presume that results of the experiments in simple "ideal" man-machine environments may be extended to more complex systems as a starting point.

The present work was carried out in a simple supervisory environment with experimentally controlled variables.

1.4 Measurements of Mental Workload

The Major difficulty with the investigation of mental workload is its inaccessibility to direct measurements. Mental events, including mental workload, are "intervening variables", in the sence that they are "middle" processes between measurable stimuli and measurable responses (Sheridan and Simpson, 1979). The "black box" nature of MWL requires the researcher to determine appropriate sets of variables as stimuli and responses (inputs and outputs), and then infer the properties of MWL from the behavior of these variables.

There are certain conditions that should be met inorder for a measure to be applicable (Ephrath in Sheridan and Simpson, 1979). These conditions are, of course, dictated by common sense. In addition, any measure of MWL should have "external validity": it should be established whether the properties of MWL could be inferred from the measure. Any measure should also have "internal validity", in the sense that it reflects some "reliable" degree of consistency - e.g., agreement between subjects (Chiles, 1977; Sheridan and Simpson, 1979).

Techniques of MWL measurement, in general, fall under four distinct categories: physiological measures, measures based on spare mental capacity hypothesis, primary task measures, and subjective measures (a comprehensive reference: Wierwille and Williger, 1978).

1.4.1 Physiological measures. The physiological method involves the scientific measurement of some physiologically related processes in human operators, such as respiration rate, heart rate, and many others. The implicit assumption is that changes in workload result in involuntary changes in some physiological processes of the human body. There exists a high degree of variability in all of these measures; and the validity of inter-relation assumption has yet to be established. However, it is a common measurement technique and well covered in literature (for a case study see Ruffel Smith, 1979).

1.4.2 Measures of spare mental capacity. The basis underlying the spare mental capacity concept is the assumption of single-channel, sampling model of human operator (Knowles, 1963; and Rolfe, 1973). The spare mental capacity is the difference between total mental capacity available and that used in performance of the primary task by the HO. The workload index in manual tracking proposed by Levison (in Moray, 1979) is based on similar assumption of fixed channel capacity of the controller.

The mental load is assumed to be proportional to the used mental capacity. The "secondary task" measures are supposed to indicate the degree of "sparecity" (unused portion) of mental capacity. In general, experiments consist of the HO working on both primary and a secondary task. There exist some controversy regarding the choice of secondary tasks, and methods of measurement which would warrant least interference with primary tasks (Rolfe, 1973; and Levine, Ogden, and Eisner, 1978).

1.4.3 Primary task measures. The primary task performance is not necessarily limited to the output of the man-machine system. It also has to do with changing strategy, planning, and behavior of HO. The basis of these measures lie in the understanding that the MWL somehow affects the performance of operators. These measures could take many forms, since there are different conditions involved in any actual or experimental situation.

1.4.4 Subjective measures. Subjective measures are based on the subjective judgements of MWL by the human operators. Subjective judgements, given that proper questions were asked, are said to be inherently (externally) valid; hence, they provide the basis on which other measures are validated.

The subjective measurements of MWL could be divided into two categories depending on the assumptions of single-dimensionality or multi-dimensionality of subjective MWL. Sheridan and Simpson (1974) have suggested the following procedures for obtaining subjective scales of MWL:

a- Single-dimensional subjective scale of MWL:

-- Category scales: similar to Cooper-Harper scale (1969).

-- Cross-modality matching: developed by Stevens (1972).

-- Thurstonian or pockilific scales: (see References: 4, 21, 22 and 24); Thurston's Law of comparative judgment is the basis of scaling technique in the present work.

-- Direct magnitude stimation: based on utility models.

b- Multi-dimensional subjective scales of MWL:

-- Separate scales for different aspect of MWL.

-- Judging conditioned differences.

-- Policy capturing and multi-attribute utility.

-- Multi-dimensional scaling.

1.5 Plan of This Thesis

1.5.1 Summary of what was done. The present work consists of subjective measurement of MWL in a simulated multi-task environment. In Section 1.3 of this chapter the author attempted to justify the need for experimentation in an ideal controlled invironment. The experimental paradigm - simulated multi-task environment - was developed by Tulga (1978) as his Sc.D. thesis work. Experimental subjects were asked to "attend" different cases of Tulga's decision-making tasks. After the subjects had attended to a pair of cases for 100 seconds each they were asked to compare the cases, and to give a subjective assessment of the relative mental workload induced by the pair. By accumulation of these paired comparisons data, response matrices were constructed whose elements were one of the five possibilities (certainly greater, probably greater, equal, probably less, certainly less) for "relative" mental workload induced by the cases. The response matrices were then transformed into the scale values of mental workload for each case by applying the comparative law of judgment (Thurstonian technique), after making some simplifying assumptions. The final data was statistically analyzed. The degrees of agreements between subjects were established, and the effects of the parameters of multi-task environment in the variation of MWL was studied.

In Chapter 2 properties of the experimental paradigm (multi-task environment) are described. Also reported are: description of the conditions of the experiments, what subjects actually did, and the procedures for the administration of subjective measurements.

Chapter 3 presents the necessary theoretical background on the scaling technique together with the procedure for the reduction of data, and the resulting subjective scales of MWL.

Included in Chapter 4 are various statistical analyses of data and their results.

Finally, conclusions and recommendations for further research are presented in Chapter 5.

1.5.2 Why Subjective Judgment? The main objective of the research was to find out what factors in a multi-task decision environment contribute to the operator's "sense" of mental workload. The experiment was not concerned with the moment to moment (transient) changes in MWL. Each experimental case was identified with a set of task variables, and the (steady state) mental workload associated with those variables was evaluated after each subject had worked on a case for a certain period of time

(100 seconds). The overall variation among cases was not exhaustive, i.e. it was impossible to determine if the experimental cases represented all the possible decision-making situations. Also, the measurement had to be administered without interference to the actual performance of the subjects. And finally, the measure had to be a valid indicator of "mental" workload.

Based upon these considerations, and in the absence of any established theory of MWL, subjective judgment as conscious experience of mental effort was decided to be the most plausible method of measurement.

Mental workload, being a man-machine "interface" phenomena, should inevitably reflect the general characteristics of both the dynamic systems and human operators. A definitive theory of mental workload should provide a semantically exact description of the physical process. The author believes that an internally valid subjective measurement of MWL and its correlation with physically observable variables of the system would provide the basis on which the "vague" concept of MWL could be translated into a more "exact" notion. By exact notion of MWL, it is meant to be the proper identification of those properties of the man-machine interface that influence the operator's sense of mental effort.

1.5.3 Why the Thurstonian technique was used.

In the beginning, the effort was focused on the application of multi-dimensional scaling techniques and appropriate computer algorithms for the purpose of constructing a multi-dimensional scale of subjective mental workload.

Preliminary experiments were conducted in order to generate data to be used as input to a multi-dimensional scaling computer program (in this case: INDISCAL, developed by Carroll). Subjects were to compare pairs of experimental cases and to judge how "dissimilar" the mental workload associated with each pair were. In other words, they had to give magnitude estimations on some scale of dissimilarity of mental workloads.

This proved to be an unwarranted demand on the subjects. The experiment failed to generate a workable set of data because the subjects consistently failed to establish the most dissimilar pair of cases - which were needed as extremums on the magnitude scale of dissimilarity of MWL. In addition, subjects also consistently failed to report the same magnitude estimation of dissimilarity (on a scale of 100) for a given pair on numerous trials.

The author believes that the following factors were responsible for the above results:

- Subjects were unable to establish an internal representation (and a frame of reference) for a MWL dissimilarity scale. Subjects confused the judgement of the dissimilarity of MWL (attribute) with the judgement of the dissimilarity of experimental cases (stimuli).

- Subjects could not establish a frame of reference for the judgement of distance of dissimilarity.

- The above, may have been the results of the difficulty of comparison. A subject had to work for 100 seconds on each case, and therefore, was unable to retain the memory of his/her previous judgments as a frame of reference.

However, the subjects felt comfortable in comparing a given pair of cases, and to judge the "relative" order of their sense of associated mental workloads. Therefore, the method of paired comparisons - with an exhaustive set of judgment categories in the form of "certainly greater", "probably greater", "equal", "probably less", and "certainly less" - was decided to be the most appropriate technique of measurement.

Then, it was found that Thurstone's Judgment scaling model and the law of comparative judgment could be employed to construct unidimensional scales of subjective MWL for the data generated by the method of paired comparisons.

CHAPTER 2

THE EXPERIMENT

2.1 Experimental Paradigm and the Man-Machine Interface

The experimental paradigm developed by Tulga (1978) simulates a multi-task decision-making situation. A multi-task decision-making situation is characterized by a number of blocks (tasks) of differing dimensions simultaneously displayed on the CRT (Figure 2.1), randomly appearing and moving to the right toward a deadline, after which they disappear. Each block is characterized by its "importance" (indicated by the height of each block) and the operator's "productivity" (the rate at which the width of a block is decreased through the action of the cursor).

The subject (human operator, HO; or decision maker) "attends" to these tasks one at a time by holding the cursor of a data-tablet to the right of the block (Figure 2.2), when thus "attended" the width of a block decreases at a constant rate. The subject is asked to maximize his/her total value gained, which is the sum of the reduction in areas of all blocks attended to by him, i.e. the total area diminished.

Blocks appear randomly with Poisson arrival and move at differing speeds towards the dealine. Height or

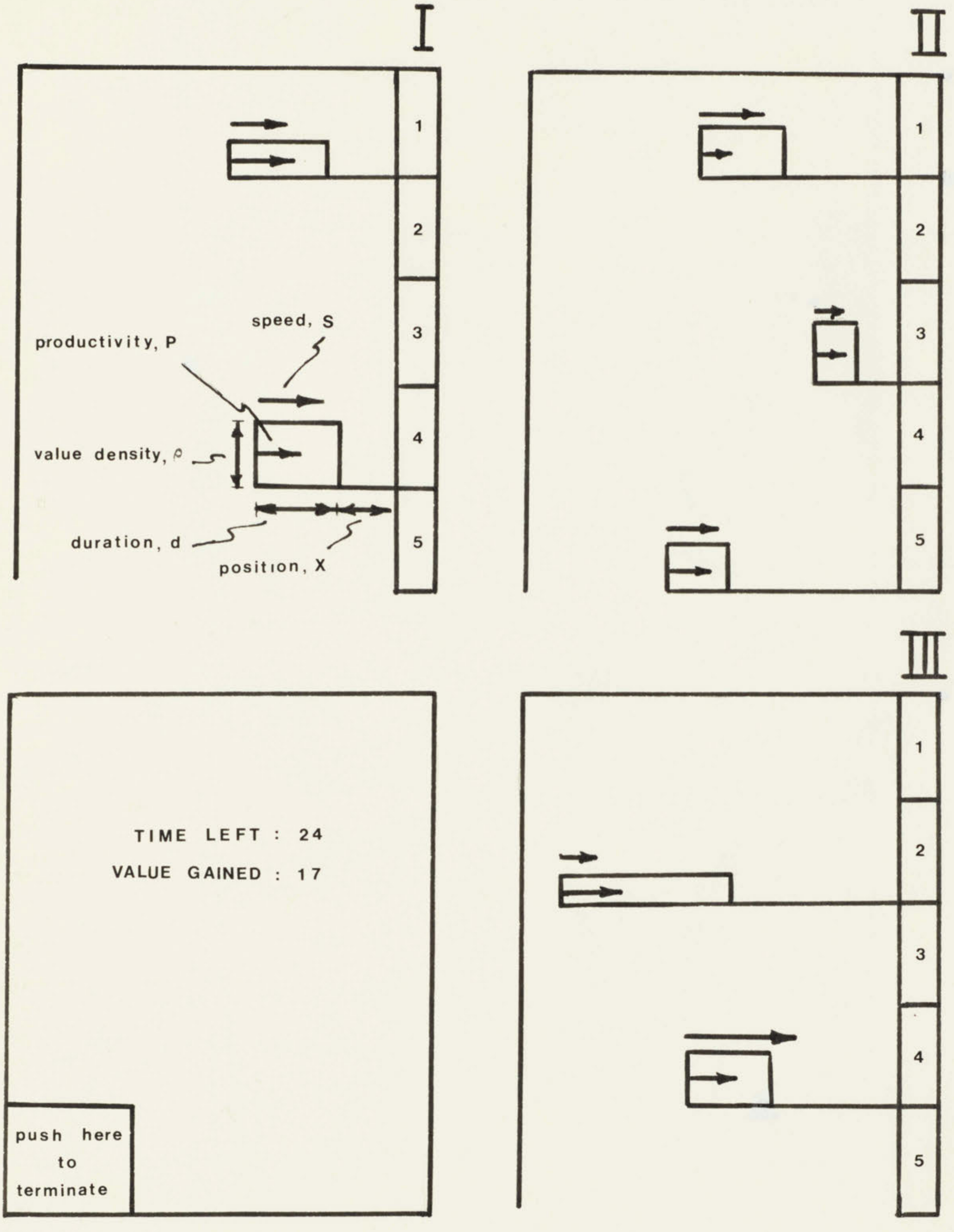


Figure 2.1 Paradigm of Dynamic Task Demands with Multiple (3) Queues.

Figure 2.2 What the Subject Sees and Does.

"importance" of a block can be thought of as the "value density" of the task which indicates the benefits accrued per unit time the HO acts on it. Value can then be earned as the time integral of value densities of tasks acted upon. The value density for each case has a rectangular probability density function of 1.0 unit/time.

The explicit parameters that differ for each case (trial) are speed, productivity, and the interarrival time of the blocks (tasks). For three levels of each, a factorial experiment requires 27 cases. Figure 2.3 displays the properties of the baseline Case 1. Properties of all the other cases with respect to Case 1 are shown in Figure 2.4.

It should be noted from Fig. 2.1 that the tasks are grouped into three different ensembles (or queues). Hence, there may be "transition time" losses (τij) for the HO when he/she transfers his/her action (moves the cursor) from i^{th} queue to the j^{th} one. Transition times do not directly enter the analyses performed later, except in the formulation of loading factor (Chapter 4).

In short, the HO monitors the arrival of different tasks, evaluates the tasks, chooses one, and acts on it before it hits the deadline.

The experiments were implemented on an INTERDATA Model-70 miniprocessor coupled with two DIABLO Series-30

BASELINE CASE : (CASE # 1)

NUMBER OF QUEUES : 3

THE PROBABILITY DENSITY FUNCTIONS (pdf) :

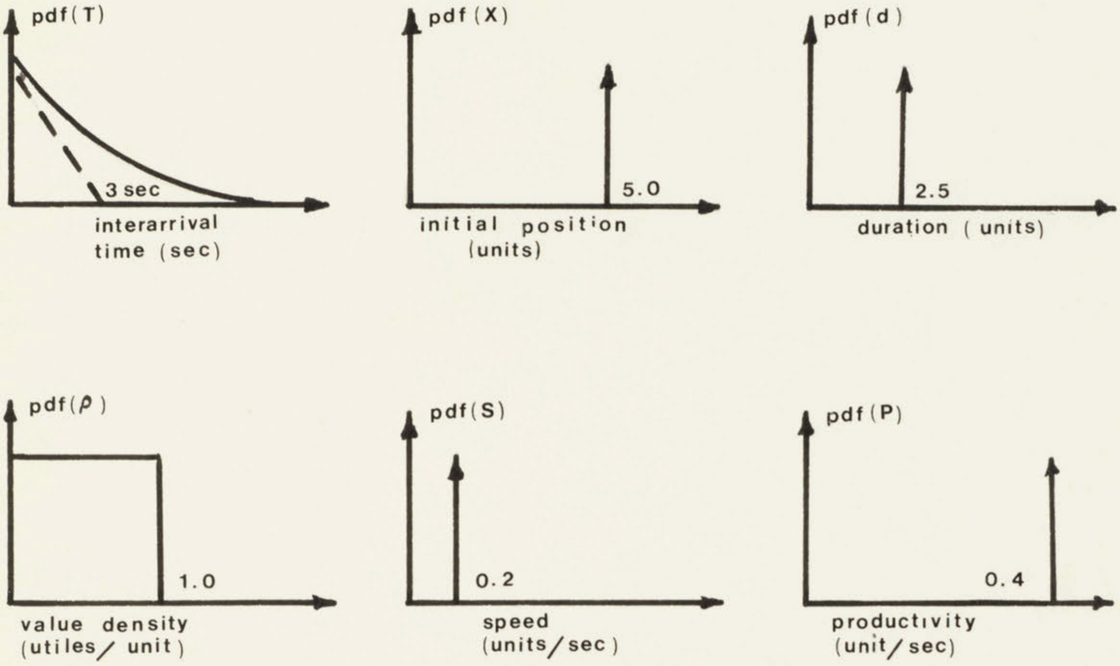


Figure 2.3 Properties of the Baseline Case.

CASES	INTERNATIONAL TIME T (seconds)	TASK SPEED S (units/sec)	PRODUCTIVITY P (Units/sec)
1	T_1 = 3	S_1 = 0.2	P_1 = 0.4
2	3	0.2	P_2 = 0.8
3	3	0.2	P_3 = 1.6
4	3	S_2 = 0.4	0.4
5	3	0.4	0.8
6	3	0.4	1.6
7	3	S_3 = 0.8	0.4
8	3	0.8	0.8
9	3	0.8	1.6
10	T_2 = 6	0.2	0.4
11	6	0.2	0.8
12	6	0.2	1.6
13	6	0.4	0.4
14	6	0.4	0.8
15	6	0.4	1.6
16	6	0.8	0.4
17	6	0.8	0.8
18	6	0.8	1.6
19	T_3 = 12	0.2	0.4
20	12	0.2	0.8
21	12	0.2	1.6
22	12	0.4	0.4
23	12	0.4	0.8
24	12	0.4	1.6
25	12	0.8	0.4
26	12	0.8	0.8
27	12	0.8	1.6

Figure 2.4 Properties of the Experimental Cases

Disk Drives, an IMLAC PDS-1 Dynamic Display CRT and SHINTRON Ecricon-250 graphic data-tablet.

2.2 Subjects and the Subjective Judgment

One female subject (operator 1), and two male subjects (operators 2 and 3) were invited to participate in the experiment. After initial stages of training each subject was asked to "attend" to multi-task decision-making cases for 100 seconds each; and then compare the cases on a pair-wise basis in order to give a subjective assessment of the relative mental workload induced by the pair.

Cases were presented in a random fashion. Care was taken to avoid any noticable order in the presentation of cases. Most of the pairs were presented more than once and some even up to five times in the course of the experiment. For each subject the experiment was carried out in a period of two to three weeks for a total of 35 to 45 hours.

The paired comparisons judgements of a subject resulted in one of the three categories:

a - The subject was certain that the mental workload induced by one of the cases was greater than the other.

b - The subject thought that probably the mental workload induced by one of the cases was greater than the other.

c - The subject was unable to make a relative judgment between the cases.

2.3 Data Recorded

By accumulation of these paired comparisons data, response matrices were constructed whose elements were one of the five possiblities (certainly greater, probably greater, equal, probably less, certainly less) for "relative" mental workload induced by each pair of cases. Due to replication, as shown in the example of Fig. 2.5a, some elements contained more than one response. The diagonal of matrices were left blank, assuming that the response of any subject when a case is compared to itself would be "c" (of the three categories a, b, and c of the last section). In addition, in order to indicate whether a case represented by a row induced more mental workload or the one represented by the column, the verbal responses were added a "+" or "-" prefix. Hence, the response matrix at this stage became a skew symmetric matrix of the example in Fig 2.5a. These verbal responses were transformed into the frequency response form by the criteria displayed in Table 2.1.

CASES	1	2	3	4
1		+a	+b +a	+b c
2	-a		c	-b +a
3	-b -a	c		+a +a
4	-b c	+b -a	-a -a	

Figure 2.5a - An Example of a Verbal Response Matrix

CASES	1	2	3	4
1		4	7	5
2	0		2	5
3	1	2		8
4	3	3	0	

Figure 2.5b - An Example of a Frequency Response Matrix

Category	Verbal Response	element ij i: row j: column	element ji j: row i: column
I: +aij	i certainly greater than j	4	0
II: +bij	i probably greater than j	3	1
III: cij	i equal to j	2	2
IV: bij	i probably less than j	1	3
V: -aij	i certainly less than j	0	4

Table 2.1 Verbal to Frequency Transformation Criteria

The above criteria translate into the following:

I: is equivalent to judging stimulus i to be greater than stimulus j four times in four trials.

II: is equivalent to judging stimulus i to be greater than j three times in four trials.

III: is equivalent to judging stimulus i to be greater than stimulus j twice in four trials.

Cases IV and V are opposites of II and I, respectively. An example of frequency response matrix is shown in Fig. 2.5b.

In Appendix A the original data in the form of frequency response are presented for the three subjects and also in the aggregate form. Appendix B presents an example of the reduction and transformation data to the final scale form, as explained in the next chapter.

CHAPTER 3

THE SCALING METHOD AND REDUCTION OF THE DATA

3.1 Introduction*

In Section 1.5.3, the method of paired comparisons was deemed to be the appropriate technique of measurement. In such a measurement the data is in the form of "proportion" of times a stimulus is observed to be greater or smaller than another stimulus. Assuming absence of any physical correlate for the psychological attribute of mental workload this question should be answered: what do these "proportions" tell about the distances between the stimuli on a psychological continuum?

To deal with a similar question, Thurndike (Reference 25) in 1910 assumed that the differences between distances of pairs of stimuli are proportional to the differences in the "unit normal deviates" corresponding to the proportions for each pair. Later, Thurstone (1927) presented a mathematical model for relating scale values of a set of stimuli to observable proportions. The "assumption" made by Thurndike could be deduced from a special case of Thurstone's model.

---

* Torgerson's "Theory and methods of scaling", has been of much help in preparation of this chapter.

The "law of comparative judgment", which is based on Thurstone's judgment scaling model, provides the necessary simplifying assumptions and the set of equations that transform the paired comparisons data into the scale values of mental workload.

In summary, Thurstone's model and Thurndike's assumption, suggest that the psychological distance between two stimuli is proportional to the normal deviate transform of the proportions of times a difference between the stimuli is noticed. As is explained in subsequent sections, an absolute index of proportion for a pair of stimuli is highest (1 or -1) when one of the stimuli is always judged greater, and it is lowest (0.5) when there is a total confusion. Therefore, the subjective scale of Thurstone is based upon the degree of confusion of the subject in the judgment of relative intensity of the psychological attribute in a pair of stimuli. A larger degree of confusion (lesser order of consistency) results in a smaller distance between a pair of stimuli.

3.2 Thurstone's Judgement Scaling Model

Thurstone postulates that for any given attribute (e.g., mental workload) of a series of stimuli (e.g., decision tasks) there exists a "psychological continuum" associated with that attribute. A subject presented with a series of stimuli would react "discriminally" with a respect to the given attribute (i.e. the subject would discriminate between the different levels of MWL associated with a series of decision tasks). The process by which the subject identifies the attribute and reacts discriminally to it is called a "discriminal process"*; and each of the "discriminal responses" associated with a discriminal process has a value on the psychological continuum associated with that attribute. It is assumed that due to the stochastic nature (noise generation, momentary fluctuation) of the organs of the human mind (sensation and cognition), i.e. of the discriminal processes, the discriminal response associated with a given attribute of a stimuli could be thought as having a frequency distribution on the psychological continuum. Furthermore, it is postulated that the frequencies with which the discriminal responses are associated have the form of a normal distribution.

In short, presenting a subject with a stimulus a large number of times results in a normal frequency distribution of a discriminal response on a psychological continuum. The standard deviation of the distribution of dis-

*It should be noted that here the "discriminal process" is taken to be the process that results in a "discriminal response": An input (stimulus) to the system (discriminal process) results in an output (discriminal response). Torgerson makes no such distinction and wherever he uses the first term, I use the second!

criminal processes, i.e. of the response to a fixed stimulus is called the "discriminal dispersion". In a normal frequency distribution the modal, mean, and the mode coincide. A possible choice for the "scale value" of the attribute on the psychological continuum is the "modal discriminal responses" (the discriminal response most often associated with the attribute for a given stimulus).

Figure 3.1 provides examples of the distribution of discriminal responses on a psychological continuum for four stimuli. Note that the modal discriminal responses s_1, s_2, s_3 and s_4 and the respective discriminal dispersion vary for each stimulus.

It is assumed that the subject cannot directly report the modal value of the discriminal responses or their dispersion on the psychological continuum. However, presumably he can judge and report relations among stimuli (e.g., paired comparisons in the form of proportions).

3.3 The Law of Comparative Judgment

The law of comparative judgment is a set of equations relating the scale values and discriminal dispersion of a set of stimuli on the psychological continuum to the proportion of times any stimulus is judged greater than the others for a given attribute.

When a pair of stimuli is presented to a subject,

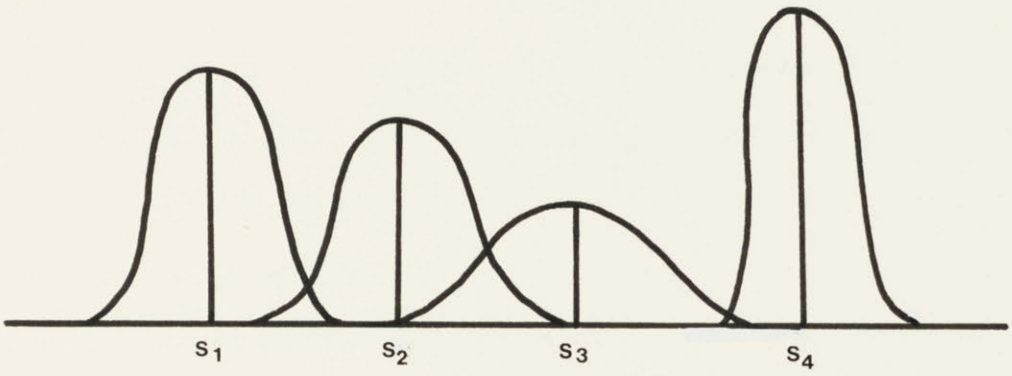


Figure 3.1 Distribution of Discriminal Responses Associated with Four Stimuli on the Psychological Continuum.

the difference between two discriminal responses (discriminal difference, $d_k - d_j$), should also form a normal distribution. The mean of this distribution is equal to the difference in scale values of the two stimuli. The standard deviation of differences is computed from

$$\sigma_{d_k - d_j} = (\sigma_j^2 + \sigma_k^2 - 2\, r_{jk}\, \sigma_j \sigma_k)^{1/2} , \qquad (3.1)$$

where r_{jk} is the correlation between momentary values of discriminal responses associated with stimuli j and k.

If there exists an overlap in the distribution of discriminal processes (Figure 3.2a), then, for a series of trials, the discriminal difference ($d_k - d_j$) could be positive or negative; or in other words, one stimulus would be judged greater than the other only some of the time.

It is possible to determine the theoretical difference ($S_k - S_j$) of scale values of a pair of stimuli from the distribution of the discriminal differences and the table of areas under the unit normal curve. Figure 3.2b, illustrates the distribution of discriminal differences. The shaded area corresponds to the proportion of times ($d_k - d_j$) is positive, or the proportion of times stimulus k is judged greater than stimulus j. The distance from the zero point to the mean of the distribution, denoted by x_{jk}, is evaluated from the table of areas under the unit normal

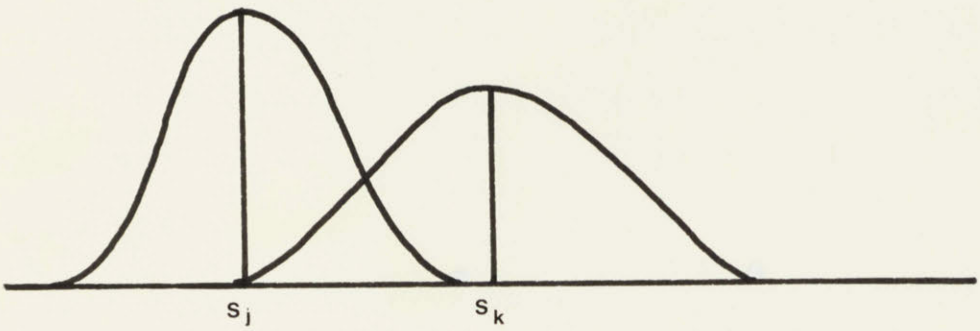


Figure 3.2a Distribution of Discriminal Responses Associated with Stimuli j and k on the Psychological Continuum.

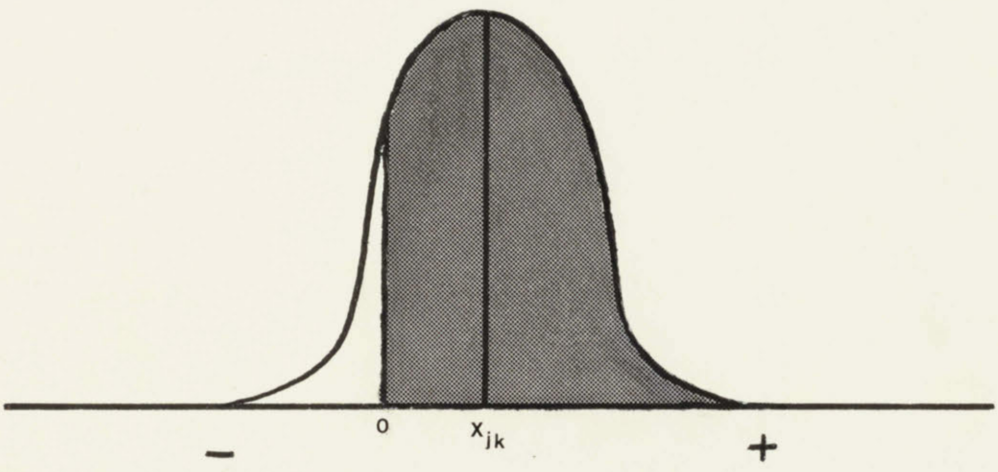


Figure 3.2b Distribution of Discriminal Differences on the Psychological Continuum. The Quantity x_{jk} is the difference in Scale Values measured in σ_{dk-dj} units.

P_{jk}	X_{jk}
0.0000005	-4.9
0.005	-2.57
0.3085	- .5
0.5	0.0
0.6915	+ 0.5
0.995	+ 2.57
0.9999995	+ 4.9

Figure 3.2c Examples of Correspondence Between Scale Differences X_{jk} and Proportion Values P_{jk}.

curve, and it is in units of the standard deviation of discriminal differences $\sigma_{d_k-d_j}$. Some examples of the distribution of discriminal differences are presented in Figure 3.2c. The difference x_{jk}, measured in $\sigma_{d_k-d_j}$, represents the difference in scale values of the two stimuli (S_k-S_j); and hence, the following holds:

$$S_k - S_j = x_{jk}\,\sigma_{d_k-d_j} \quad . \qquad (3.2)$$

Combining Equations 3.1 and 3.2, the fundamental equation of the law of comparative judgement results:

$$S_k - S_j = x_{jk}\,(\sigma_j^2 + \sigma_k^2 - 2\,r_{jk}\,\sigma_j\sigma_k)^{1/2} \qquad (3.3)$$

In general, for n stimuli, there are n scale values, n discriminal dispersion, and n(n-1)/2 independent correlations which are unknown. Against these, there are only n(n-1)/2 observable equations corresponding to the independently observable proportions. In order to decrease the number of unknowns some simplifying assumptions are necessary.

3.4 Replication Classes and Simplifying Conditions

According to Torgerson (Ref. 23), the experiments fall into three classes of data replications:

Class I) Replication over trials within a single individual: the normal distribution of discriminal differences for a pair of stimuli results from the presentation of the pair in many trials.

Class II) Replication over individuals, each pair of stimuli being compared once by each individual: the normal distribution of discriminal differences results from the interchangability of subjects.

Class III) Replication over both individuals and trials: the normal distribution of discriminal differences results from the interchangability of subjects and many trials.

Note that even though the law of comparative judgment has an identical form for these three classes, the fundamental assumption concerning the formation of distribution of the discriminal processes is interpreted differently for each class.

The simplifying conditions which result in sets of workable equations fall into three categories (Torgerson, 1958):

Condition A) Assumption of constant correlation (r_{jk} = constant).

Condition B) Assumption of equal correlations and small differences in discriminal dispersion (r_{jk} = r, constant; $\sigma_k - \sigma_j$ = d: small).

Condition C) Assumption of constant variance of discriminal differences ($\sigma_k = \sigma_j$ = c: constant).

Of these, Condition C results in the simplest workable equation. Later, the significance of these classifications and their effect on the results of this experiment will be discussed.

3.5 A Workable Set of Equations

In Condition C it is assumed that the standard deviation of discriminal differences is constant and the same for all pairs of stimuli. This assumption results in the following set of equations:

$$S_k - S_j = C\, x_{jk}\,, \qquad (3.4)$$

where C denotes the constant discriminal dispersion ($\sigma_{d_k - d_j}$ = C). Here, C is taken to be equal to unity; and since the discriminal dispersion is the unit of measurement, this sets the unit for the final scales.

Equation 3.4 holds exactly if the values of x_{jk} are theoretically exact. However, the experiment provides only the estimated value x_{jk}' instead of true x_{jk}. A "least square" solution for the estimated scale value of the stimuli results in the following:

$$S_k' = \frac{1}{n} \sum_{j=1}^{n} x_{jk}' + \frac{1}{n} \sum_{j=1}^{n} S_j' \quad (k = 1,2 \ldots .n) \qquad (3.5)$$

For convenience, the origin of the scale is set at the mean of the estimated scale values:

$$\frac{1}{n} \sum_{j=1}^{n} S_j' = 0 \qquad (3.6)$$

Hence, the estimated scale values of stimuli are computed from

$$S_k' = \frac{1}{n} \sum_{j=1}^{n} X_{jk}' \quad (k = 1,2, \ldots n) . \qquad (3.7)$$

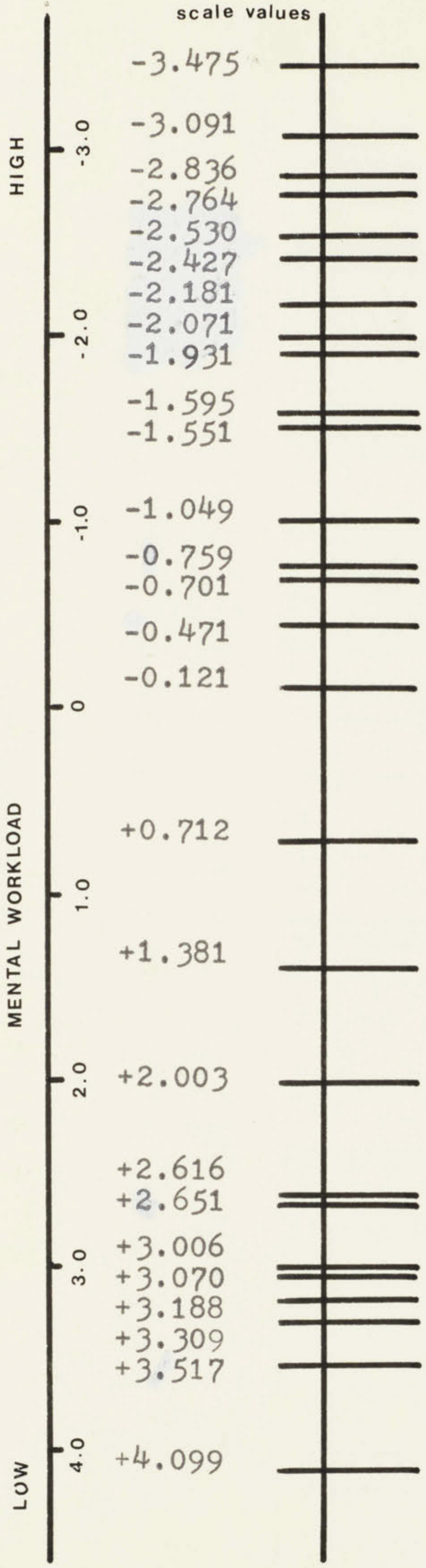
It should be noted that the scale values derived through use of the law of comparative judgment locate the stimuli on the psychological continuum with respect to one another only; and since the method itself cannot determine an absolute zero point, it must be chosen arbitrarily.

3.6 Reduction of the Data

The data generated in the experiment had to be transformed into a set of workable information. The response data for each subject was compiled into a frequency response matrix F whose rows and columns represented the stimuli and the elements were the "number of times the stimulus k was judged greater than stimulus j for a given attribute" (Section 2.5). Then, the main response matrix F was transformed into a proportion matrix P whose elements were the "proportion of times the stimulus k was judged greater than stimulus j for a given attribute". The basic transformation matrix X was constructed from matrix P. The element x'_{jk} is the unit normal deviate corresponding to the element P'_{jk}, and it is obtained from the table of areas under the unit normal curve. The values of x'_{jk} are unbounded for unity values of proportion P'_{jk}. A value of 5 was taken to be a reasonable estimate for the unbounded x'_{jk}'s (i.e. placing the zero point in Figure 3.2b, 5 standard deviations away from the mean of the distribution of discriminal differences). However, any other value - if it is reasonably large - could be chosen; and it would somehow set the distance between two extreme points on the scale. The reader is recommended to compare the scale of Fig. 3.4 (based on the value of 5 units of s.d.) to that of Fig. 3.5 (based on the value of 50 unites of s.d.).

3.7 Scales of Subjective Mental Workload

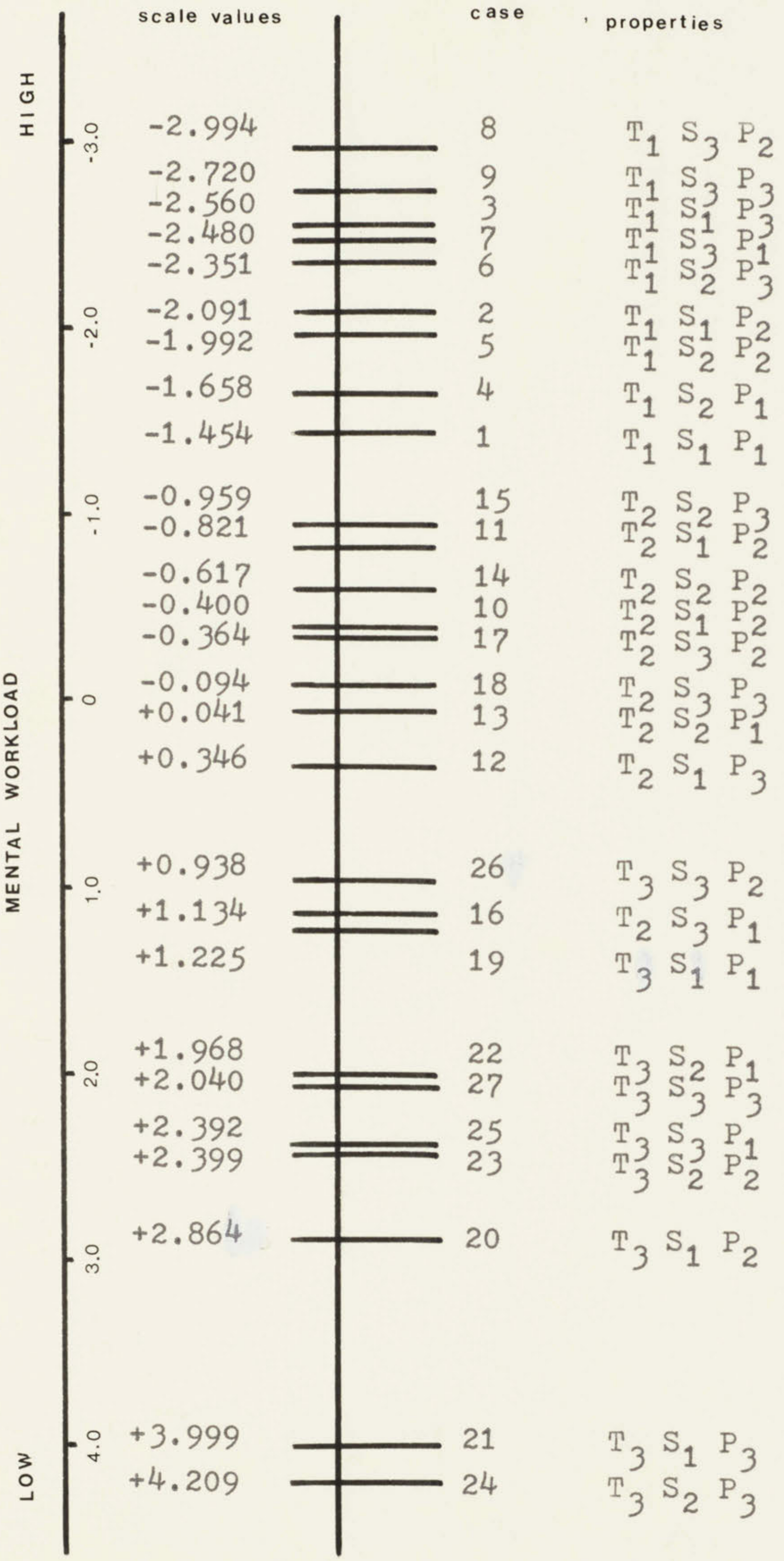
By substituting the values of x'_{jk} in Equation 3.7 subjective scale values of mental workload for each subject was constructed. These scales are presented in



scale values	case	properties
-3.475	2	T_1 S_1 P_2
-3.091	1	T_1 S_1 P_1
-2.836	3	T_1 S_1 P_3
-2.764	9	T_1 S_3 P_3
-2.530	4	T_1 S_2 P_1
-2.427	6	T_1 S_2 P_3
-2.181	7	T_1 S_3 P_1
-2.071	8	T_1 S_3 P_2
-1.931	5	T_1 S_2 P_2
-1.595	11	T_2 S_1 P_2
-1.551	10	T_2 S_1 P_1
-1.049	12	T_2 S_1 P_3
-0.759	14	T_2 S_2 P_2
-0.701	13	T_2 S_2 P_1
-0.471	15	T_2 S_2 P_3
-0.121	18	T_2 S_3 P_3
+0.712	17	T_2 S_3 P_2
+1.381	19	T_3 S_1 P_1
+2.003	26	T_3 S_3 P_2
+2.616	23	T_3 S_2 P_2
+2.651	16	T_2 S_3 P_1
+3.006	20	T_3 S_1 P_2
+3.070	22	T_3 S_2 P_1
+3.188	24	T_3 S_2 P_3
+3.309	27	T_3 S_3 P_3
+3.517	21	T_3 S_1 P_3
+4.099	25	T_3 S_3 P_1

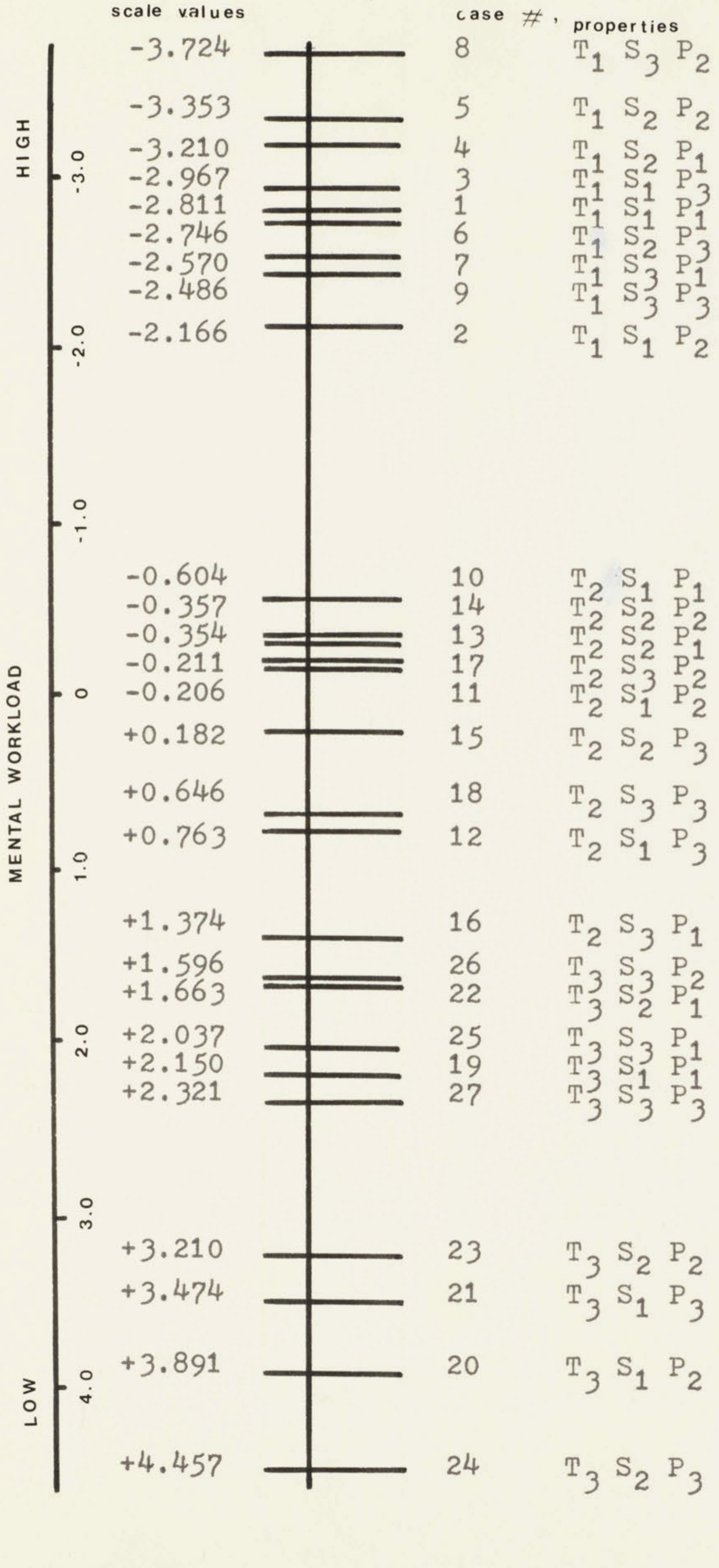
T: interarrival time
S: speed
P: productivity

Figure 3.3a Scale of Subjective Mental Workload for Operator 1



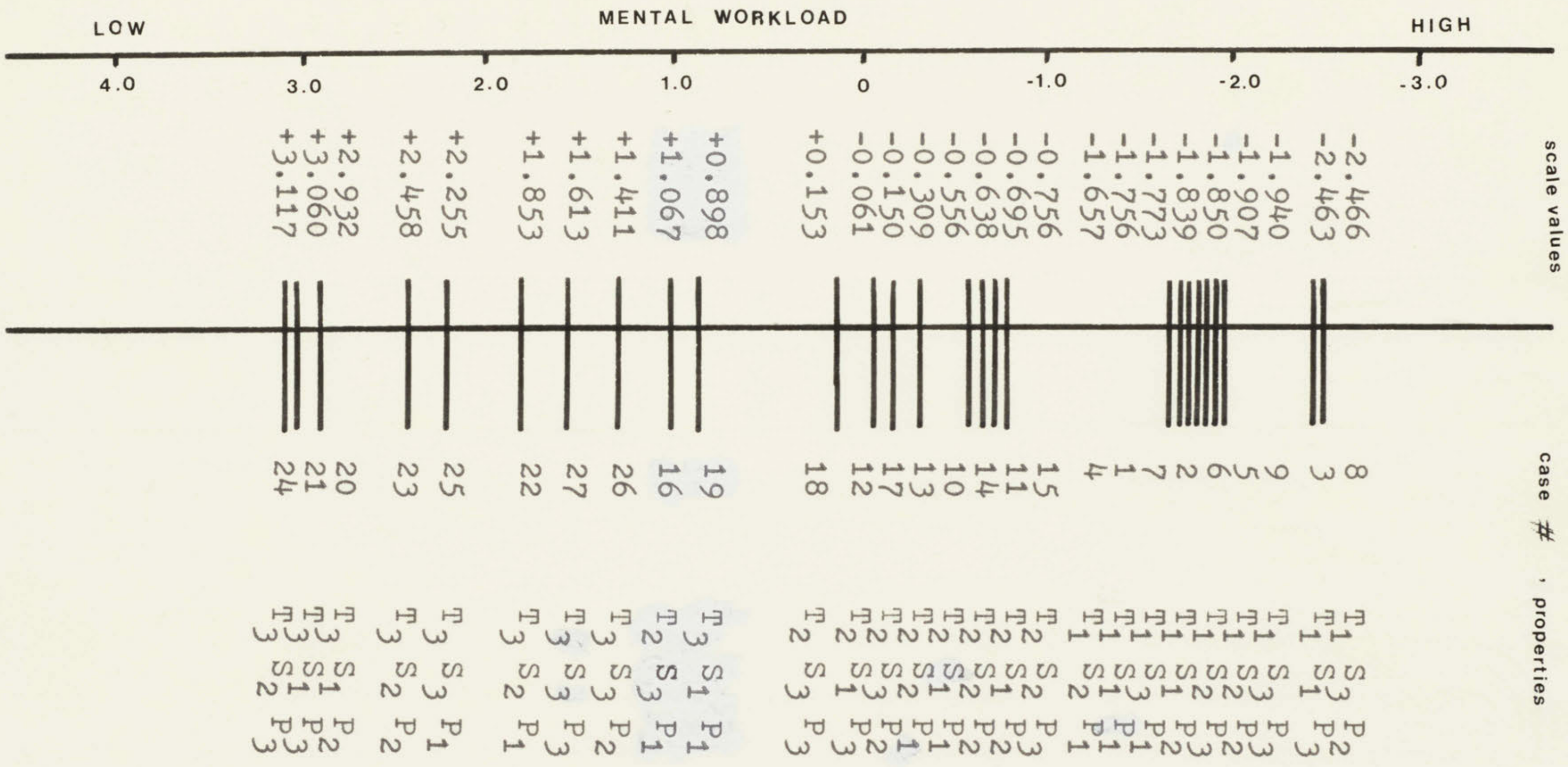
T: interarrival time
S: speed
P: productivity

Figure 3.3b Scale of Subjective Mental Workload for Operator 2



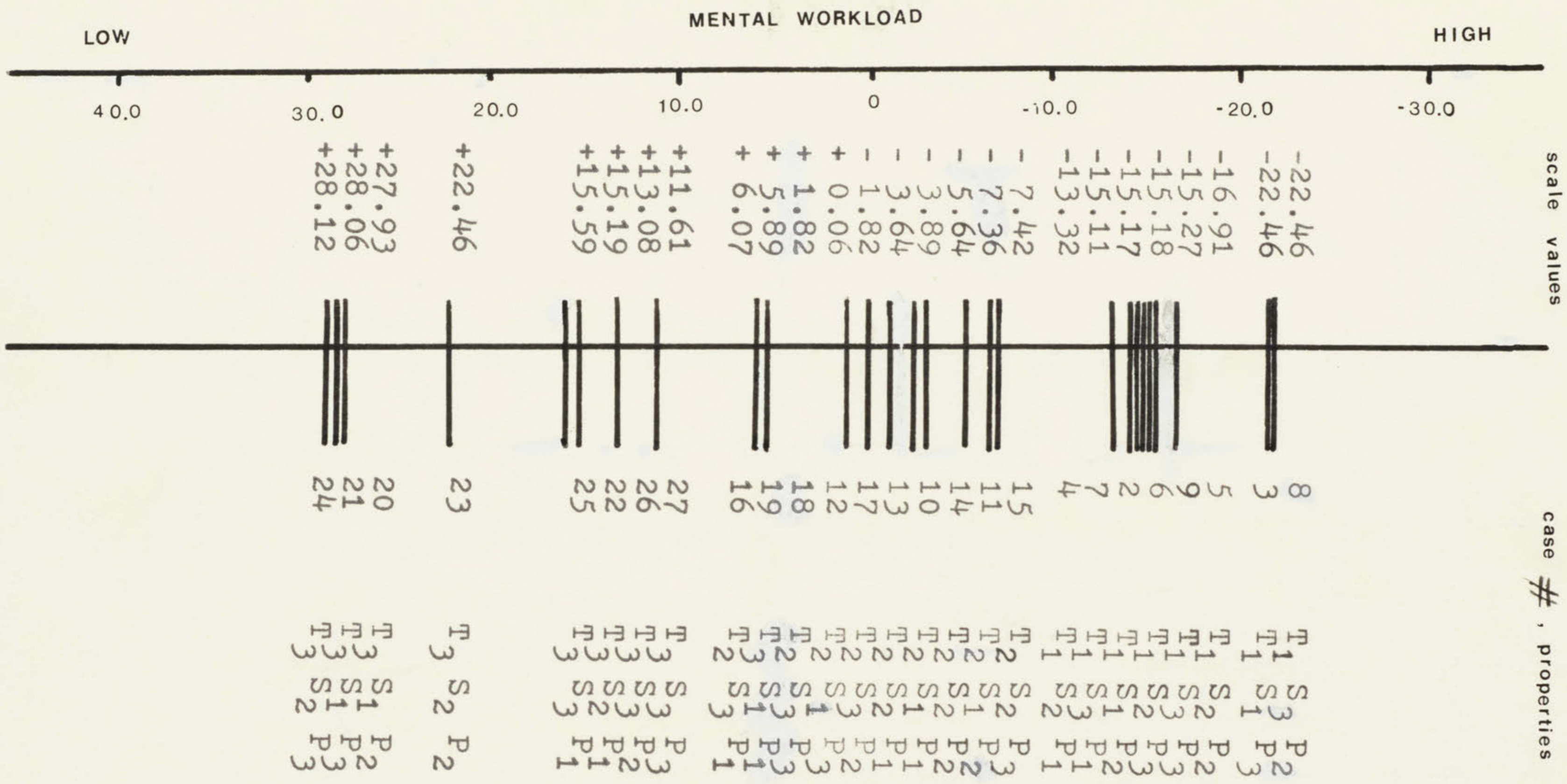
T: interarrival time
S: speed
P: prodoctivity

Figure 3.3c Scale of Subjective Mental Workload for Operator 3



T: interarrival time
S: speed
P: productivity

Figure 3.4 Scale of Subjective Mental Workload for Aggregate of Data (X_{jk} = 5.0 for P_{jk} = 1)



T: interarrival time
S: speed
P: productivity

Figure 3.5 Scale of Subjective Mental Workload for Aggregate of Data (X_{jk}= 50 for P_{jk}= 1)

Figures 3.3a, b, and c. Figure 3.4 shows the aggregate scale which results from combining the responses of all subjects together.

CHAPTER 4: ANALYSES OF DATA: RESULTS AND INTERPRETATION

4.1 Introduction

In this chapter three independent techniques of analysis have been employed:

I) Analysis of variance, which tests the significance of the effects of different variables of decision-task environment on the variation of MWL.

II) Analysis of agreement among the subjects, which tests how well subjects agree among themselves in their judgments.

III) Analysis of transitivity, which provides a basis upon which the consistencies of the judgments of subjects are compared.

A four-way analysis of variance which was performed on the rank order data indicates that the interarrival time T has the most significant effect on the variation of MWL. This could be observed from the examination of the scales of subjective MWL in Figures 3.3a, 3.3b, 3.3c and 3.4. Based upon this finding, the analyses mentioned above were performed on data at each level of interarrival time, and also on the overall data including all levels of interarrival time.

Subsequent sections discuss the analyses and their results.

4.2 Analysis of Variance

4.2.1 Introduction

The experimental method of paired comparisons and subsequent application of the Law of Comparative Judgment produces a Thurstonian scale of MWL as described in Chapter 3. The MWL for each decision case is represented by a single scale value of its associated MWL. The Thurstonian scale, therefore, provides the rank order of cases with respect to MWL as well as an interval scale (Table 4.1).

A four-way analysis of variance was performed on the aggregate rank order data to test the significance of the effects of different operators, interarrival times, task speeds, and operator productivities (O,T,S, and P's respectively) on the variation of MWL. As mentioned before, the analysis showed that effect of interarrival time T was highly significant. Subsequently, three-factor analyses were performed separately for each level of interarrival times.

4.2.2 Four-way Analysis of Variance on Aggregate Data

The results of the analysis of variance on the aggregate rank order data are presented in Table 4.2. The final form of rank order data (Tables 4.1a,b) includes no replication, resulting in zero residual variance. There-

Table 4.1a

ORIGINAL RANK ORDER DATA FROM THURSTONIAN SCALE

	T_1									T_2									T_3								
	S_1			S_2			S_3			S_1			S_2			S_3			S_1			S_2			S_3		
	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3
O_1	2	1	3	5	9	6	7	8	4	11	10	12	13	14	15	21	17	16	18	22	26	23	20	24	27	19	25
O_2	9	6	3	8	7	5	4	1	2	13	11	17	16	12	10	19	14	15	20	25	26	21	24	27	23	18	22
O_3	5	9	4	3	2	6	7	1	8	10	14	17	12	11	15	18	13	16	21	26	25	20	24	27	22	19	23

Note: Each entry is the rank order of MWL within each operator (subject)

T = Interarrival Time

S = Speed of the Tasks

P = Productivity

O = Operators

Table 4.1b

Original Data of Table 4.1a Put Into Coded Form by Subtracting 14 From Each Value

	T_1									T_2									T_3								
	S_1			S_2			S_3			S_1			S_2			S_3			S_1			S_2			S_3		
	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3
O_1	-12	-13	-11	-9	-5	-8	-7	-6	-10	-3	-4	-2	-1	0	1	7	3	2	4	8	12	9	6	10	13	5	11
O_2	-5	-8	-11	-6	-7	-9	-10	-13	-12	-1	-3	3	2	-2	-4	5	0	1	6	11	12	7	10	13	9	4	8
O_3	-9	-5	-10	-11	-12	-8	-7	-13	-6	-4	-0	3	-2	-3	1	4	-1	2	7	12	11	6	10	13	8	5	9

TABLE 4.2

Results of Analysis of Variance for all Levels of Interarrival Time

Nature of effect	Source	Sums of squares	d.f.	Variance Estimate	Significance Level %
Main factors	T *	4303.19	2	2151.60	1
	S	9.85	2	4.93	-
	P *	32.67	2	16.34	15
	O	0	2	0.00	-
Interaction between Pair of factors	TS *	84.74	4	21.19	1
	TP *	55.92	4	13.98	5
	TO	1.03	4	0.26	-
	SP *	93.26	4	23.32	1
	SO	93.71	4	23.43	1
	PO	24.44	4	6.11	-
Interaction between Triples of factors	SPO	36.07	8	4.51	-
	TPO	38.08	8	4.76	-
	TSO	32.31	8	4.10	-
	TSP *	51.04	8	6.38	17†
Interaction of all factors	TSPO	57.19	16	3.57	
	Residual	0	0	0	
TOTAL		4914	80		

*: Significant effects

†: Computed by linear interpolation

fore, the variance estimate due to the highest order interaction (TSPO) provides the basis for the significant tests.

The effects of higher order interaction are discussed first, since the variance estimate of higher order interactions were progressively lumped into the residual to test the effects of the lower order interactions. Thus, the main factor effects are the last ones discussed. Note that only those interactions which were found to be statistically significant are discussed here. Throughout the analysis the "null hypothesis" maintains that none of the interactions or main factors have any significant effect on the variation of MWL.

a. Second Order Interaction. As Table 4.2 indicates, the effect of the TSP interaction is statistically significant (at %17 level). The high degree of complexity involved renders it impossible to interpret the mechanism of interaction; however, the MWL plots of TSP interaction in Fig. 4.1 provide the impression that the effect of interarrival time T could be very significant.

b. First Order Interactions. Interaction TS, SP and SO are found to have highly significant effects (significant at % 1 level). The effect of the TP interaction is significant as well, but to a lesser degree (significant at %5 level). Interpretation of these interactions will

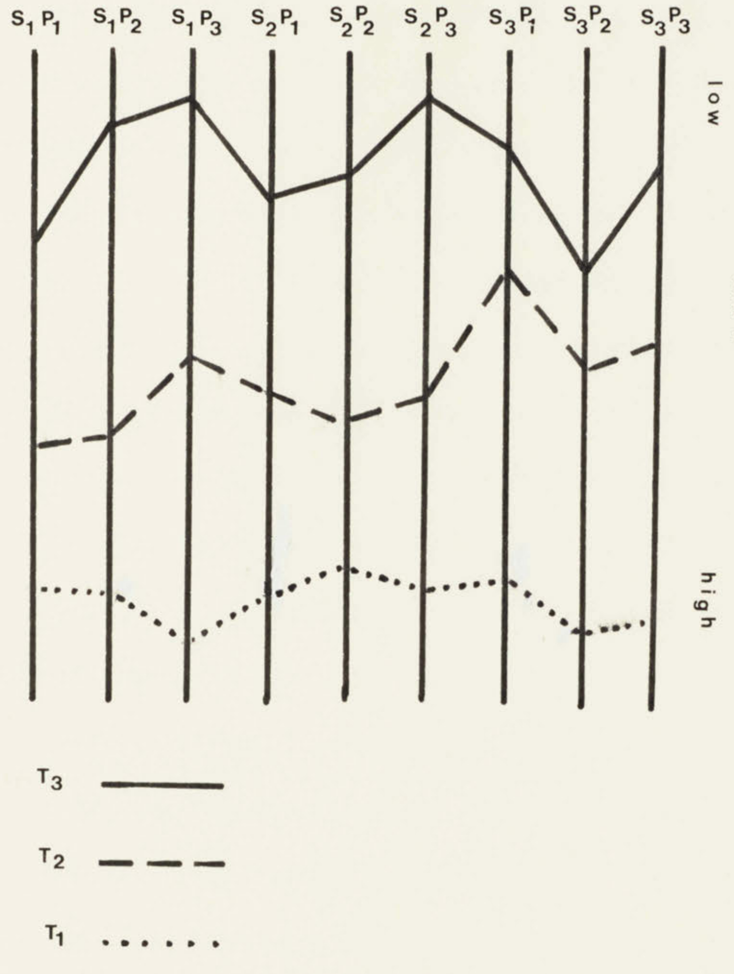


Figure 4.1 Effect of "Interarrival time x Speed X Productivity" Interaction on the Variation of Mental Workload.

be taken up later separately in the discussion results for different levels of interarrival times.

c. Main Factor Effects. Interarrival time (T) has the most significant effect in the variation of MWL, as shown in Table 4.2. Note that the rank order data has, almost completely, been divided into three zones of high, medium, and low levels of MWL corresponding to low, medium, and high levels of interarrival time, respectively.

A closer inspection of experimental cases shows that the decision-task environments for the three selected levels of T are very unsimilar in terms of performance required from the operators. The decision-task environment is discussed separately for each level of T.

I. Short Interarrival Time (T_1 = 3 seconds). This level of T is characterized by high concentration of tasks on the display. In Table 4.3, column M lists the number of tasks on the display at any instant, which for T_1 level varies between 25 and 6.25 depending on the value of task speed S. In the same table column N lists the total number of tasks to be processed in 100 seconds. For T_1 level, there are 100 tasks to be processed in 100 seconds. These cases require a higher level of monitoring and decision making in comparison to the cases at the T_2 and T_3 level. In Table 4.3, column J and/or K indicate that many tasks go unattended.

II) Medium Interarrival Time (T_2 = 6 seconds). At this level of T new tasks arrive every 6 seconds. Table 4.3, column M indicates that there are 12.5 to 3.12 tasks on the display at every instant depending on the level of task speed S. There are a total of 50 tasks to be processed every 100 seconds (Table 4.3, column N). Hence, there is less concentration on tasks on the screen compared to T_1 level, and consequently a lower level of monitoring/ decision-making involved. There are few cases that may be completely processed before the tasks hit the deadline (cases 12,15,18 from Table 4.3, columns J and K).

III) Long Interarrival Time (T_3 = 12 seconds). At this level of T new tasks arrive every 12 seconds. This level of T is characterized by little or no monitoring/ decision-making at all; and in some cases there is free time between the arrival of tasks. This level of T is operationally different from the others. The operator does not have to worry about monitoring and decision-making, since there are not too many tasks on the display. Hence, he/she is more concerned with the prompt action required to minimize the manual transfer time in moving the hand from one task to the other. Also, the existance of free time contrasts sharply with the continuous mental occupation

TABLE 4.3

Properties of Decision Cases

	Case	Inter-arrival time $1/\lambda$ or T	Task Speed S	Productivity P	sec^{-1} λ or $1/T$	sec^{-1} $\frac{s}{x}$	sec $\frac{x}{s}$	sec $\frac{d}{p}$	Primary factor	Load Factor LF*	L<1
$T_1S_1P_1$	1	3	.2	.4	.33	.04	25	6.25	T	6.25	
$T_1S_1P_2$	2	3	.2	.8	.33	.04	25	3.12	T	3.12	
$T_1S_1P_3$	3	3	.2	1.6	.33	.04	25	1.56	T	1.56	
$T_1S_2P_1$	4	3	.4	.4	.33	.08	12.5	6.25	T	6.25	
$T_1S_2P_2$	5	3	.4	.8	.33	.08	12.5	3.12	T	3.12	
$T_1S_2P_3$	6	3	.4	1.6	.33	.08	12.5	1.56	T	1.56	
$T_1S_3P_1$	7	3	.8	.4	.33	.16	6.25	6.25	T	6.25	
$T_1S_3P_2$	8	3	.8	.8	.33	.16	6.25	3.12	T	3.12	
$T_1S_3P_3$	9	3	.8	1.6	.33	.16	6.25	1.56	T	1.56	
$T_2S_1P_1$	10	6	.2	.4	.17	.04	25	6.25	T	3.12	
$T_2S_1P_2$	11	6	.2	.8	.17	.04	25	3.12	T	1.56	
$T_2S_1P_3$	12	6	.2	1.6	.17	.04	25	1.56	T	.78	*
$T_2S_2P_1$	13	6	.4	.4	.17	.08	12.5	6.25	T	3.12	
$T_2S_2P_2$	14	6	.4	.8	.17	.08	12.5	3.12	T	1.56	
$T_2S_2P_3$	15	6	.4	1.6	.17	.08	12.5	1.56	T	.78	*
$T_2S_3P_1$	16	6	.8	.4	.17	.16	6.25	6.25	T	3.12	
$T_2S_3P_2$	17	6	.8	.8	.17	.16	6.25	3.12	T	1.56	
$T_2S_3P_3$	18	6	.8	1.6	.17	.16	6.25	1.56	T	.78	*
$T_3S_1P_1$	19	12	.2	.4	.08	.04	25	6.25	T	1.56	
$T_3S_1P_2$	20	12	.2	.8	.08	.04	25	3.12	T	.78	*
$T_3S_1P_3$	21	12	.2	1.6	.08	.04	25	1.56	T	.39	*
$T_3S_2P_1$	22	12	.4	.8	.08	.08	12.5	6.25	T	1.56	
$T_3S_2P_2$	23	12	.4	.8	.08	.08	12.5	3.12	T	.78	*
$T_3S_2P_3$	24	12	.4	1.6	.08	.08	12.5	1.56	T	.39	*
$T_3S_3P_1$	25	12	.8	.4	.08	.16	6.25	6.25	X	3.00	
$T_3S_3P_2$	26	12	.8	.8	.08	.16	6.25	3.12	X	1.5	
$T_3S_3P_3$	27	12	.8	1.6	.08	.16	6.25	1.56	X	.74	*
COLUMN		A	B	C	D	E	F	G	I	J	K

$$* \quad LF = \left(\frac{d}{p} + \tau\right) \sum_{i=1}^{I=3} \max\left(\frac{1}{T}, \frac{s}{x}\right) \qquad d = 2.5,\ \tau = 0,\ x = 5.0$$

Table 4.3 (Cont.)

CASE					$0 \le 1$		$q > 1$
1	25	25	100	4		1.33	*
2	25	25	100	8		2.67	*
3	25	25	100	16		5.33	*
4	25	12.5	100	2		.67	
5	25	12.5	100	4		1.33	*
6	25	12.5	100	8		2.67	*
7	25	6.25	100	1	*	.33	
8	25	6.25	100	2		.67	
9	25	6.25	100	4		1.33	*
10	12.5	12.5	50	4		1.33	*
11	12.5	12.5	50	8		2.67	*
12	12.5	12.5	50	16		5.33	*
13	12.5	6.25	50	2		.67	
14	12.5	6.25	50	4		1.33	*
15	12.5	6.25	50	8		2.67	*
16	12.5	3.12	50	1	*	.33	
17	12.5	3.12	50	2		.67	
18	12.5	3.12	50	4		1.33	*
19	6.25	6.25	25	4		1.33	*
20	6.25	6.25	25	8		2.67	*
21	6.25	6.25	25	16		5.33	*
22	6.25	3.12	25	2		.67	
23	6.25	3.12	25	4		1.33	*
24	6.25	3.12	25	8		2.67	*
25	6.25	1.56	25	1	*	.33	
26	6.25	1.56	25	2		.67	
27	6.25	1.56	25	4		1.33	*
	L	M	N	O	P	Q	R

TABLE 4.3 (Cont.)

CASE			T>1
1	.48	.16	
2	.96	.32	
3	1.92	.64	
4	.48	.16	
5	.96	.32	
6	1.92	.53	
7	.48	.16	
8	.96	.32	
9	1.92	.64	
10	.96	.32	
11	1.92	.64	
12	3.85	1.28	*
13	.96	.32	
14	1.92	.64	
15	3.85	1.28	*
16	.96	.32	
17	1.92	.64	
18	3.85	1.28	*
19	1.92	.32	
20	3.85	1.28	*
21	7.70	2.57	*
22	1.92	.32	
23	3.85	1.28	*
24	7.70	2.57	*
25	1.92	.32	
26	3.85	1.28	*
27	7.70	2.57	*
	S	T	U

Table 4.3

Notes

I: Primary factor dominant in Load Factor equation

J: Load Factor

L: Total number of tasks to be processed in 25 seconds

M: Number of tasks present on the display at each instant

N: Total number of tasks to be processed in 100 seconds

O: Number of tasks completed from the moment a series of tasks appear to the time they hit the dead line

S: If S<1 new tasks appear before completion of any task

T: If T<1 new tasks appear before the completion of previous series of tasks

LF: Load Factor $L = \left(\frac{d}{p} + \tau\right) \sum_{i=1}^{I=3} \max\left(\frac{1}{T}, \frac{S}{x}\right)$

$d = 2.5$

$\tau = 0$

$x = 5.0$

$$\text{Load Factor} = \frac{\text{average time required to do a task}}{\text{average time affordable to do a task}}$$

at T_1 and T_2 level.

In conclusion, it should be kept in mind that the different levels of interarrival time represent fundamentally different decision environments; therefore, fundamentally different operator response behavior should be expected for these environments.

4.2.3 Analysis of Variance for Short Interarrival Time

Three-way analysis of variance was performed for the short interarrival time (T_1 = 3 seconds). Results are presented in Table 4.4.

Analysis shows that none of the main factors or interactions has any significant effect on the variation of MWL; except for the SO (subject-operator) interaction.

a. First Order Interactions. The effect of first order interaction SO is significant at the %17 level at short interarrival time T_1. The SO interaction, plotted in Figure 4.2, suggests a very dissimilar perception of speed for each operator in terms of its contribution to MWL. The attempt is made below, to interpret the results in the light of the author's observation of the operators' working behavior during the experiment:

I. Operator 1. The operator 1 was observed to be occupied with the task of monitoring and making decisions (as to which task to be attended next); and was less responsive to the prompt action which was required in mini-

TABLE 4.4

Results of Analysis of Variance at Short Inter-arrival Time (T = 3 seconds)

Name of Effect	Source	Sum of Squares	d.f.	Variance Estimate	Signif. Level
Main factor	S	6.00	2	3.00	-
	P	4.67	2	2.34	-
	O	0	2	0	-
Interaction between Pairs of Factors	SP	14.66	4	3.67	-
	SO *	76.00	4	19.00	17 +
	PO	12.89	4	3.22	-
Interaction of all Factors	SPO	65.78	8	8.22	-
	Residual	0	0	0	-
	Total	180.0	26		

* Significant effects

\+ By linear interpolation

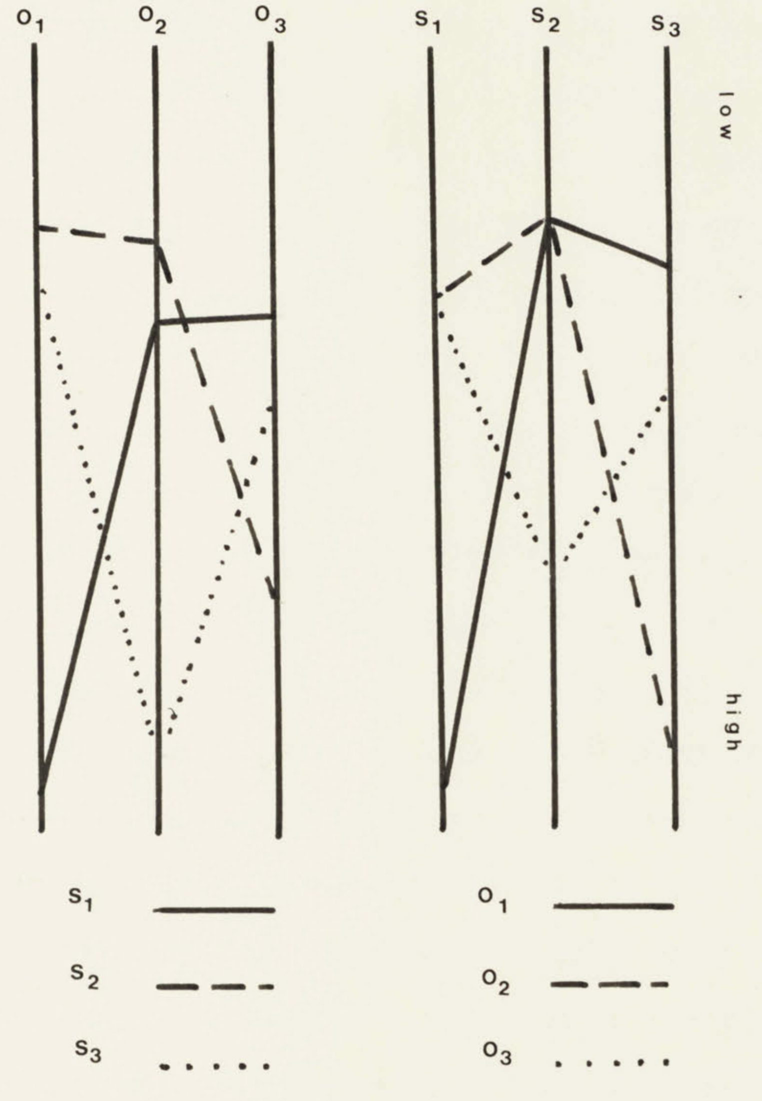


Figure 4.2 Effect of "Operator X Speed" Interaction on the Variation of Mental Workload at Short Inter-arrival time.

mizing the transfer time of the hand. A lower speed for the tasks corresponds to a longer time for them to be on the display, and therefore, to a higher concentration of tasks on the screen. For Operator 1 the lowest level of speed corresponded with the highest level of MWL, apparently due to a higher emphasis on monitoring and decision making. Higher levels of speed decrease the burden of monitoring to some extent, even though they require prompter action because of the shorter deadline for the tasks. Operator 1 was less concerned with the prompt action requirement and apparently did not act promptly anyway. Hence, higher levels of speed for the tasks resulted in low level of MWL for Operator 1. Note that S_3 speed level corresponds to 6.25 tasks on the display at any instant compared to 25 and 12.5 tasks for S_1 and S_3 speed levels.

II. Operator 2. During the experiment Operator 2 was observed to be concerned more with the prompt action requirement and less with monitoring/decision-making. The emphasis on prompt action is highest for the high speed level due to the shorter time available for the completion of each task. This partially explains the high MWL associated with high speed level for Operator 2. Note that at S_3 level of speed the available time for each task to be attended is 6.25 seconds, compared to 25 and 12.5 seconds for S_1 and S_2 levels.

III. Operator 3. During the experiment Operator 3 exhibited equal concern for both prompt action and monitoring/decision-making. The fact that Operator 3 viewed the medium level of task speed to be associated with high MWL may simply indicate that this speed level somehow culminated in the most difficult combination of prompt action and monitoring/decision-making.

4.2.4 Analysis of Variance for Medium Interarrival Time

Three-way analysis of variance was performed for the medium level of interarrival time (T_2 = 6 seconds). Results are presented in Table 4.5. Medium level of interarrival time is characterized by less abundance of tasks on the display, and therefore, less monitoring/decision-making (compared to the short T). There even exist situations (if productivity P is high) where the operators could complete all the tasks with enough prompt action.

Analysis indicates that the effect of first order interaction SP is probably significant at 18% level. Productivity P contributes to the variation of MWL, but its effect is not very significant. In contrast to the low level of interarrival time it is observed that there is no significant interaction of operators and other variables. In addition, a main factor, namely S, is found to

TABLE 4.5

Results of Analysis of Variance at Medium Interarrival Time (T = 6.0 seconds)

Name of Effect	Source	Sum of Square	d.f.	Variance Estimate	Signif. Level
Main Factor	S *	60.67	2	30.34	1
	P *	17.56	2	8.78	21 +
	O	0	2	0	
Interaction between Pairs of Factors	SP *	32.44	4	8.11	18 +
	SO	18.44	4	4.61	-
	PO	21.33	4	5.33	-
Interaction of all Factors	SPO	29.56	8	3.70	-
	Residual	0	0	0	-
	Total	180.0	26		

* Significant effects

\+ Computed by linear interpolation

have the most contribution to the MWL.

a. First Order Interactions. The effect of SP interaction is plotted in Fig. 4.3. If the effect of this interaction was not significant, then the S_1, S_2 and S_3 curves of Fig. 4.3b should have been similar. It is observed that the MWL is much lower for $T_2S_3P_1$ case compated to $T_2S_3P_3$. Table 4.3, column O indicates that for the case $T_2S_3P_1$ (#16), a task is barely completed before it hits the deadline; in addition, new tasks arrive almost at the same time the previous tasks hit the deadline (T = 6 seconds, $\frac{x}{s}$ = 6.25 seconds). Therefore, $T_2S_3P_1$ is a case with uncommon characteristics (from the point of view of the operators), due to certain combination of values of T, S and P. Mental work load associated with this case is very low because of the monotonous action required from the operators.

The only other major variation due to SP interaction appears to be at S_1 level. The MWL is much lower for the case $T_2S_1P_3$ than those of $T_2S_1P_1$ and $T_2S_1P_2$. Table 4.3 columns K and T indicate that the operators may be able to successfully complete all tasks, and may even have some free times. This is, of course, not true of the other cases.

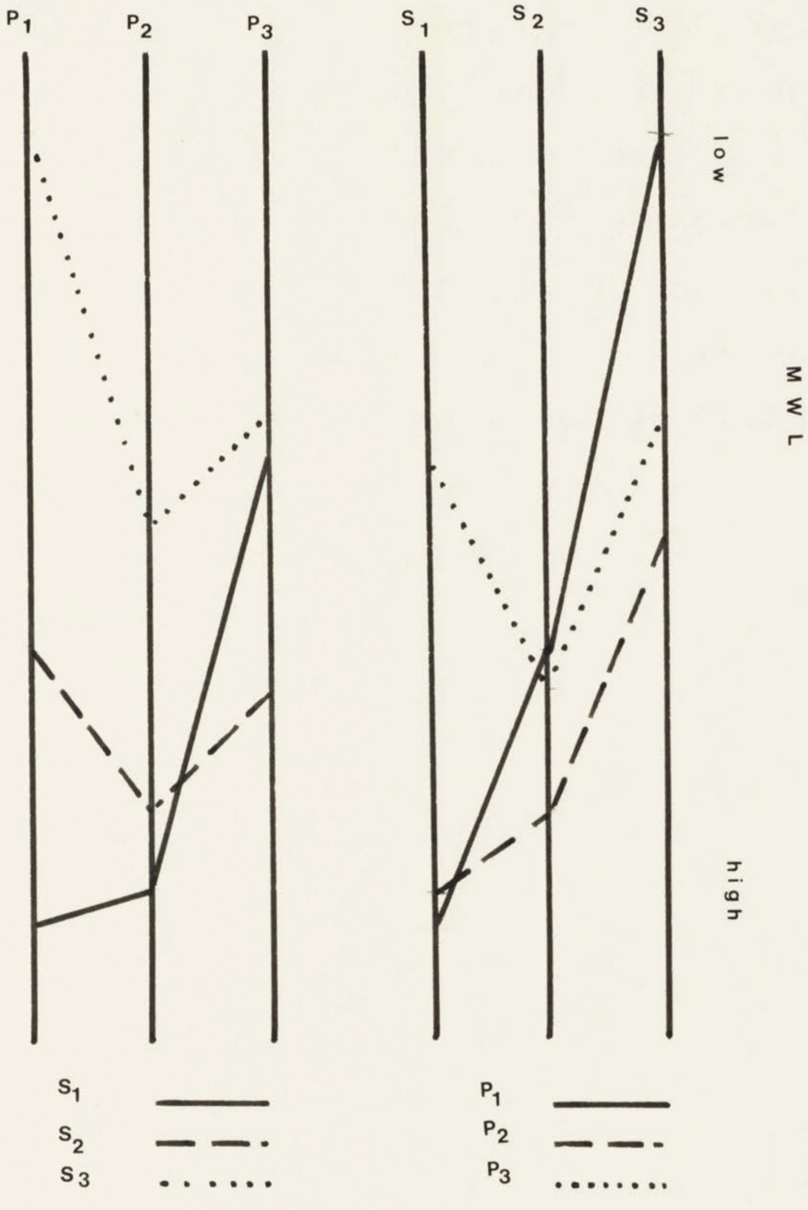


Figure 4.3 Effect of "Speed X Productivity" Interaction on the Variation of Mental Workload at Medium Interarrival Time.

b. Main Factor Effects. Among the main factors task speed S, has the most significant effect in the variation of MWL. A higher level of speed corresponds to a lower level of MWL. At S_1 and S_2 levels of task speeds, there are 12.5 and 6.25 tasks on the display at each instant (from Table 4.3, column M). Table 4.4b showes that there is not much variation MWL between S_1 and S_2 levels. For both of these speed levels the operators should maintain some degree of monitoring/decision-making. At S_3 level, however, there are only 3.12 tasks present on the display at any instant, and the burden of monitoring/decision making is substantially decreased.

4.2.5 Analysis of Variance for Long Interarrival Time

Three-way analysis of variance was performed for the long interarrival time (T_3 = 12 seconds). Results are presented in Table 4.6. This level of interarrival time is characterized by very low concentration of tasks on the display. It is noted, from Table 4.3, column M, that the instantaneous number of tasks varies between 6.25 to 1.56 tasks. Therefore, these cases require very little monitoring/decision-making. More important, in most cases operators have free times between the arrivals of new tasks (between zero and 6 seconds).

TABLE 4.6

Results of Analysis of Variance at Long Interarrival Times (T = 12.0 seconds)

Name of Effect	Source	Sum of Square	d.f.	Variance Estimate	Signif. Level
Main Factor	S	8.22	2	4.11	-
	P *	54.89	2	27.45	5
	O	0	2	0	-
Interaction between Pairs of Factors	SP *	64.65	4	17.41	1
	SO *	21.78	4	5.45	10
	PO *	13.78	4	3.45	16 +
Interaction of all Factors	SPO	11.68	8	1.46	-
	Residual	0	0	0	-
	Total	180.0	26		

* Significant effects

\+ Computed by linear interpolation

Another important aspect of this level of T is that contrary to all the previous cases there are instances where new tasks do not arrive until the old tasks have hit the deadline (this is true for the high level of speed S_3). It is noted that for these special cases the value of s/x will be the primary variable - rather than 1/T - in the equation of the Load Factor (see the bottom of Table 4.3).

As the analysis shows, of the first order interaction SP has the most significant effect. Interactions SO and PO have significant effects, but to a lesser degree. Productivity P is the most important main factor contributing to the variation of MWL.

a. <u>First Order Interactions</u>

I) <u>Speed-Productivity Interactions</u>. The effect of SP interaction in T_3 level is shown in Fig. 4.4. Since the most significant main factor effect is due to productivity P, in the absence of SP interactions similar curves should be expected for different P levels (Fig. 4.4). Curves for P_2 and P_3 are relatively similar, and totally differ in shape from the curve of P_1. Table 4.3, column K indicates that for productivity levels of P_2 and P_3 operators may be able to complete all tasks successfully (except for the case $T_3S_3P_2$) and have free times as well;

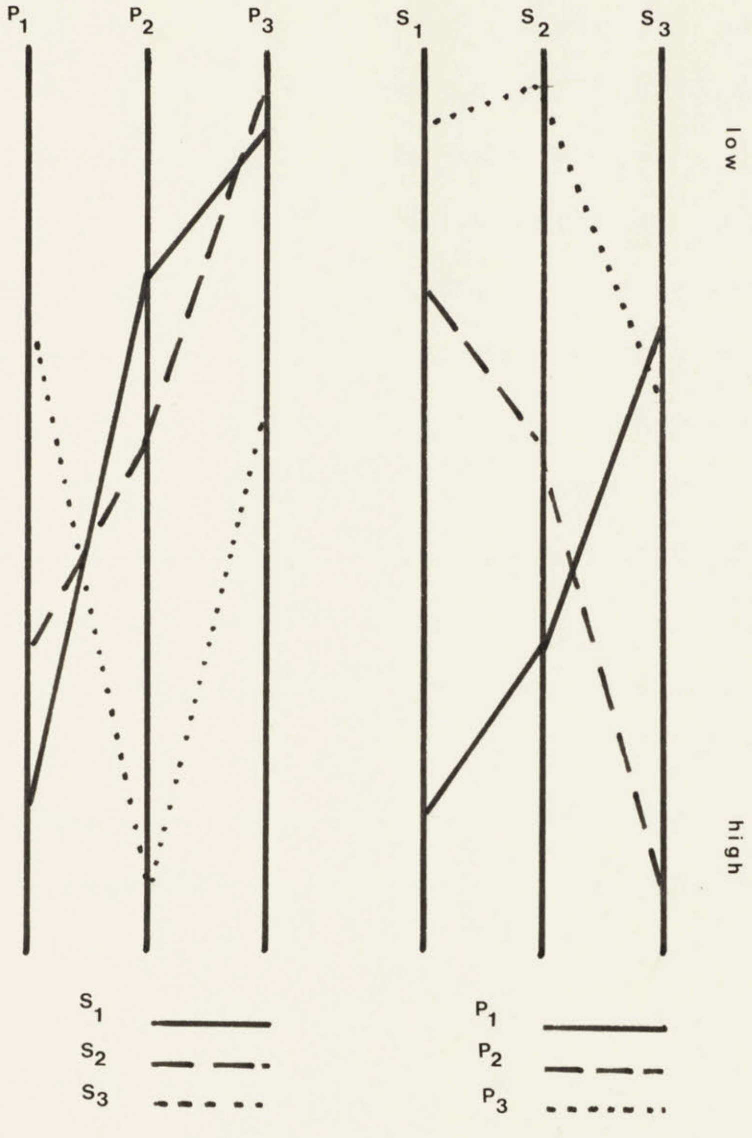


Figure 4.4 Effect of "Speed X Productivity" Interaction on the Variation of Mental Workload at Long Interarrival Time.

whereas at P_1 level not all of the tasks are completed, and there are no free times except for $T_3S_3P_1$ which explains its relative low level of associated MWL.

In conclusion, the SP interaction is significant mostly due to the fact that different levels of productivity create fundamentally different decision-task environments.

II) Operator-Speed Interaction. The effect of OS interaction in T_3 level is shown in Fig. 4.5. It is observed that the perception of the effects of task speeds for Operator 1 differs fundamentally from those of the other two operators. Higher speed provides less time for the completion of each task, hence, it requires prompter action from the operators. Operator 1, being insensitive to this requirement (see Section 5.1.2-a-I), feels less MWL at this level of speed. Note that there are also fewer tasks on the display at S_3 (Table 4.3, Column M). Operators 2 and 3 feel more MWL at S_3 level, due to their sensitivity to prompt action requirement.

Comparison of the operator behavior at low level of interarrival time (Section 4.2.2-a) and the present case shows relative consistency in the response pattern of each operator.

III) Operator-Productivity Interaction. The effect of this interaction is shown in Fig. 4.6. Again, it is

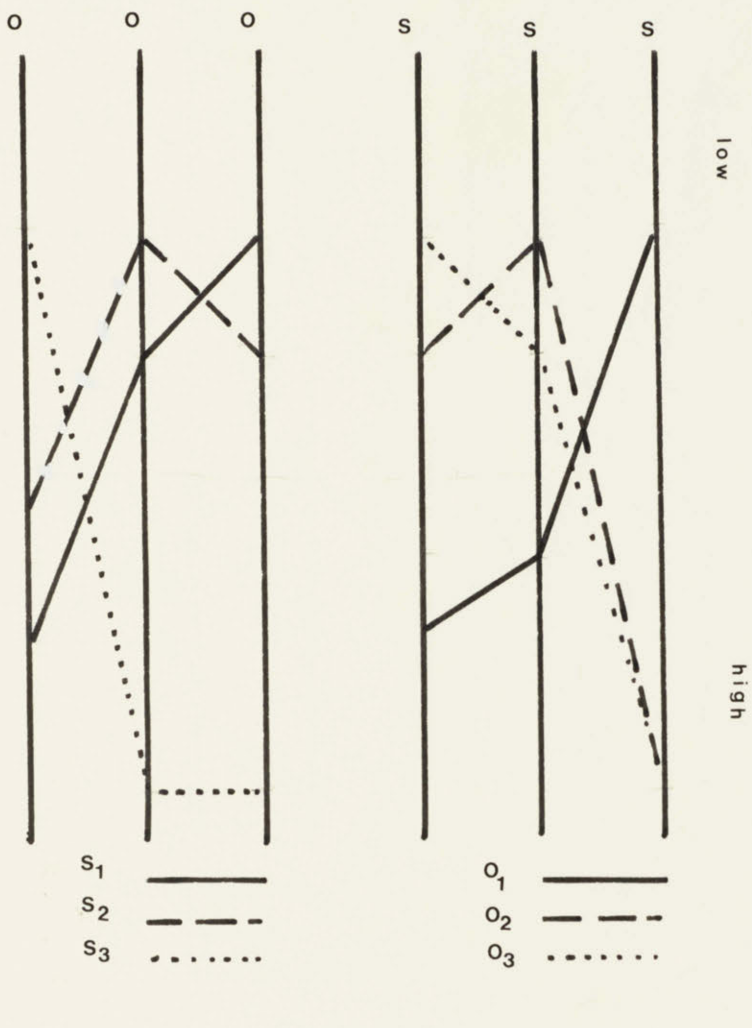


Fig. 4.5 Effect of "Operator X Speed" Interaction on the Variation of Mental Workload at Long Interarrival time.

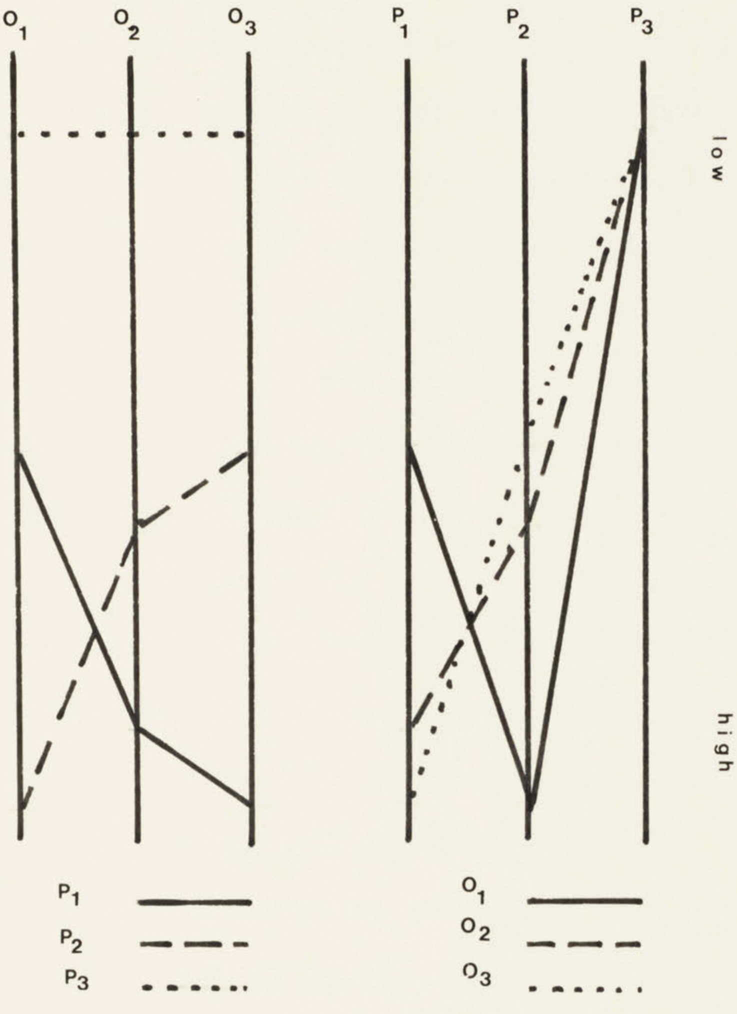


Fig. 4.6 Effect of "Operator X Productivity" Interaction on the Variation of Mental Workload at Long Interarrival time.

observed that the response patterns of Operator 1 is fundamentally different than those of the other two operators. Operator 1 feels highest MWL at the P_2 level. At this level productivity is not high enough to prohibit any monitoring/decision-making, and it is not low enough to give plenty of time for monitoring of very few tasks (monitoring/decision-making is done while a task is being attended). The emphasis of Operator 1 on monitoring rather than prompt action may be a partial explanation of the response.

Comparison of the results of this section with that of Section 4.2.2-a reveals a general similarity in MWL perception by Operators 2 and 3. Difference between the response patterns of Operator 1 and the other two operators is partially explained in terms of operational requirements of the decision environment.

b. Main Factor Effects. Among the main factors productivity P had the most significant effect in the variation of MWL. The cumulative scores of MWL with respect to productivity levels is shown in Table 4.7. Lowest productivity level (P_1) is associated with the highest level of MWL. As mentioned before, there are no free times at this level of P. Higher level of productivity (P_2) is associated with the lowest level of MWL. At this level of P the operators have the longest free times.

P_1	P_2	P_3
-11	-7	18

Table 4.7 Total Ranks of Mental Workload at Long Interarrival Time for each Level of Productivity.

4.3 Degrees of Agreement Among the Subjects

4.3.1 Introduction

A proper measure of agreement among the subjects is the "Coefficient of Concordance" denoted by W. The statistical significance of W is tested by calculating Snedecor's distribution for F. Another index of agreement which holds only for Pairs of subjects is the Spearman's Rank Correlation Coefficient denoted by R, and its significance is tested by computing the Student's t ratio. Appendix C presents the formulas for W and R and their respective significance testing methods.

These coefficients were computed both for the overall rank ordered data and for each level of interarrival time T, and their significance was tested.

4.3.2 Results

The following conclusions are made:

i) Table 4.8 indicates that there exists a highly significant overall degree of agreement among subjects (W=.94); however, the same is not true if each level of T is considered separately. This simply means that the highly significant overall degree of agreement is the result of the effect of the interarrival time on the variation of MWL, which separates the data into three regions (of high, medium, and low) MWL's.

Table 4.8

Degrees of Agreement Among Subjects

	Coeff. of Concordance	rank-diff. correl.	Spearman's rank Corr. Coeff. (R) Pair of Subjects (operators)		
	W	r'	O_1O_2	O_1O_3	O_2O_3
Overall Ranked Data	0.94	0.91	0.91	0.89	0.94
Level of Sig.	%0.5		%0.5	%0.5	%0.5
T_1 Level	0.14	0.033	-0.01	-0.29	0.18
Level of Signif.	*		*	%30-%20	*
T_2 Level	0.61	0.69	0.31	0.54	0.58
Level of Signif.	%10-%5		%20-%10	%10-%5	%5-%2.5
T_3 Level	0.74	0.83	0.61	0.47	0.96
Level of Signif.	%2.5-%1		%5-%2.5	%10-%5	%0.5

ii) The agreement among subject becomes more significant for long interarrival times, i.e. the orders of the operator's judgment become more similar at longer T's. If it is assumed that class III of the replication of data holds (see Section 3.4) - where both trials and subjects could be interchanged - we may conclude that the subject is less confused in his mind with respect to the ordering of stimuli at lower levels of MWL. At shorter values of T where MWL is very high there is not significant agreement among the subjects.

iii) Operators 2 and 3 are consistently in better agreement with each other than with operator 1. This holds true for every level of MWL. It is worth remembering from the section on analysis of variance that internal models of environment for operators 2 and 3 fundamentally differ from that of operator 1.

In light of the above results, a higher degree of reliability is placed upon the ordering of cases in lower MWL (longer interarrival times) region. It is interesting to note that in Fig. 3.4 the distance between cases becomes longer for longer interarrival times.

4.4 Transitivity and Consistency of Responses

4.4.1 Introduction

The method of paired comparisons is an appropriate technique for judgment of stimuli with multidimensional attributes. A clear indication of multidimensionality of an attribute is the intransivity of the responses of the subject. As an example of intransivity, a stimulus A may be preferred to B, and B to C; and at the same time C may be preferred to A. Examples of transivity and intransivity for triads (triplets) of stimuli are illustrated in Figure 4.7. Intransitivity is the result of inconsistency in subject's choices or preferences. In the following, transitivity and consistency (and for that matter intransitivity and inconsistency) are used interchangably when referring to the judgment of the subjects. In the following, two underlying models of attribute that result in inconsistency of responses are discussed:

a. Multidimensional attribute model. One reason for the apparent intransitivity could be that pairs of stimuli are judged on the basis of different dimensions of a multidimensional psychological attribute. For example, apple A may be preferred to apple B because it is bigger; apple B may be preferred to apple C because apple B has a better color, and apple C may be preferred to apple A because it tastes better. In this example goodness of an apple (a multidimensional attribute) consists of three independent dimensions of bigness, color and taste. If a set of stimuli possesses a multidimensional attribute with psychologically

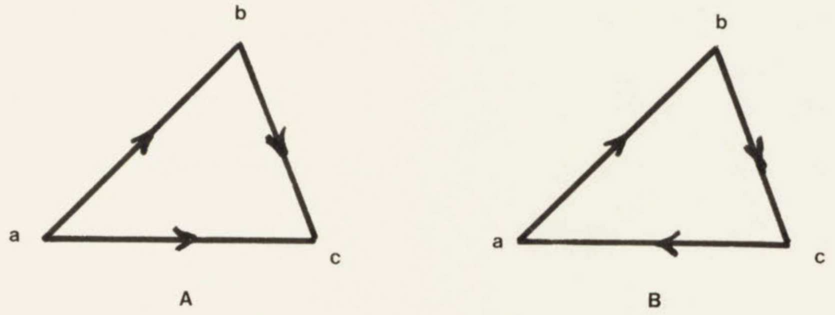


Figure 4.7 Diagram for the Transitivity of a Triad of Stimuli

Footnote: A: Transitivity resulting from consistency in choice

B: Intransitivity resulting from inconsistency in choice

(arrow indicates the direction of choice or preference) differentiable dimensions, then paired comparisons judgments of stimuli would be expected to yield highly inconsistent choices (in forms of intransitive triads of Fig. 4.7).

b. Stochastic unidimensional model of attribute. Judgment of pairs of stimuli with respect to some unidimensional attribute could, also, result in inconsistent judgments. Considering Thurstone's judgment model, because of the noise generation and momentary fluctuations in human perception, a subject's discriminal response to a stimulus may take any value (within a normal distribution) on a psychological continuum (see Section 3.2). Presenting to a subject three stimuli (a pair at a time) with small differences between their scale values - such that their associated distributions of discriminal responses overlap - may result in inconsistent judgments in the form of intransitive triads. Figure 4.8 demonstrates this situation. In an initial hypothetical judgment trial, stimuli A and B result in discriminal responses a_1 and b_1 where $a_1 > b_1$; in the second trial stimuli B and C result in the discriminal responses b_2 and c_2 where $b_2 > c_2$; and finally in the third trial stimulus A and C result in discriminal responses a_3 and c_3 where $c_3 > a_3$. Thus, the triplets of stimuli A, B, and C result in an intransitive triads.

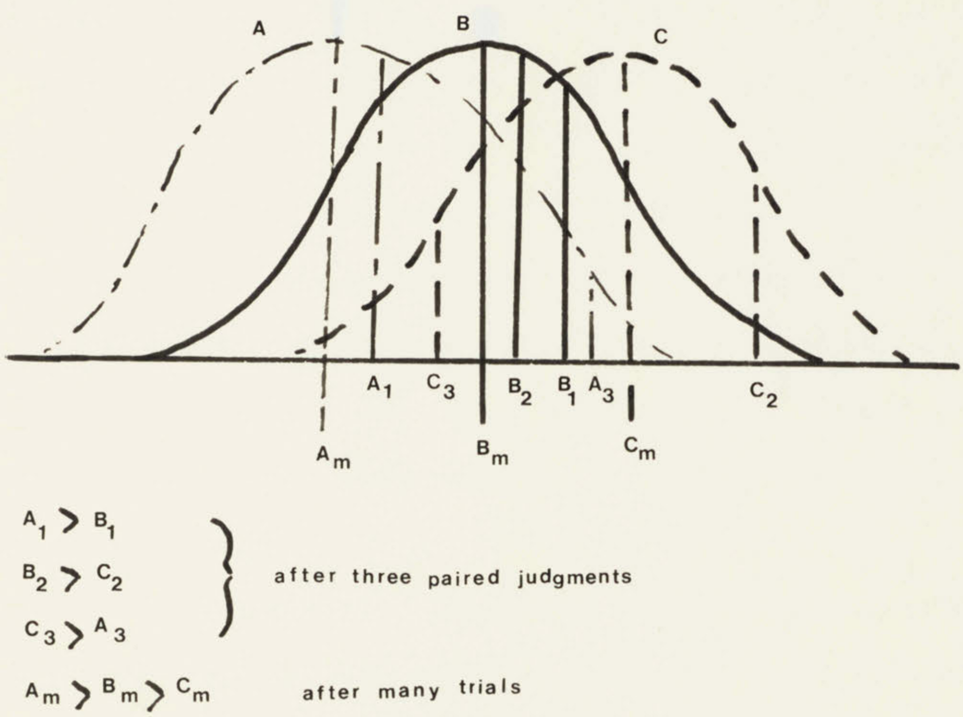


Figure 4.8 An Example of Inconsistency in Judgment of a Unidimensional Attribute.

If the distances between the theoretical scale values s_a, s_b and s_c are much larger than the "just noticable difference"*, then after many trials the modal discriminal responses a_m, b_m, and c_m should form a transitive triad; or, in other words, a set of consistent choices should emerge after many trials. However, if the distances between the theoretical scale values s_a, s_b, and s_c are equal or less than the "just noticable difference", then after many trials the triplets of the modal discriminal responses a_m, b_m, and c_m will not necessarily form a transitive triad; or, in other words, after many trials, a set of consistent choices will not emerge.

In summary, judgments of stimuli with multidimensional attributes, based on the method of paired comparisons, would result in responses with some degree of inconsistency (many intransitive triads). For stimuli with a unidimensional attribute, where normal distribution model holds, after many trials two possibilities emerge: 1) consistency increases if the distances between theoretical scale values are large, ii) inconsistency remains if the distances between theoretical scale values are equal or smaller than the "just noticeable difference".

*One possible definition: a difference that could be noticed by the resolving power of the subject.

4.4.2 Coefficient of consistency and the number of inconsistent triads.

The consistency of the responses was studied by establishing the number of inconsistent triads (triplets) of stimuli in the data. The analysis is performed on the data of the frequency or proportion matrices, and as such does not assume that Thurstone's judgment model holds valid. It should be noted that the concepts of consistency and transivity are applied to the data in the final form and are not concerned with single trials. In Appendix C an outline of the analysis procedure is given. The results of the analysis are presented in Table 5.9. The "number of inconsistent triads" is denoted by d. The "coefficient of consistency", K, is a convenient index of consistency; however, a proper technique for testing the statistical significance of K has yet to be developed.

It should be noted that there exists no apparent correlation between the results of Tables 4.8 and 4.9; and that the analyses of Table 4.8 was performed on the final rank order data for which Thurstone's judgment model was assumed to be valid.

In Table 4.9, the computed results are also presented for the overall data which is the aggregate of responses of all subject (Figure A4 in Appendix A).

	For all 27 Cases	Short inter-arrival time T_1	Medium inter-arrival time T_2	Long inter-arrival time T_2	
Oper-ator 1	100	15.5	5	12.4	d
	.87	.48	.83	.58	k
Oper-ator 2	93.7	22.0	14.8	7	d
	.88	.25	.5	.76	k
Oper-ator 3	134.2	19.2	19.1	12.2	d
	.83	.35	.36	.59	k
Aggregate Responses	44	26.2	7.5	5.25	d
	.94	.11	.75	.82	k

d: number of inconsistent triads

k: coefficient of consistency

Transivity and Consistency of Responses

TABLE 4.9

Each subject demonstrates a varying degree of consistency at different levels of T. Operators 2 and 3 become more inconsistent at lower levels of T. Operator 1 is most consistent at intermediate levels of T. Here, again, it is observed that operator 1 behaves differently from the other two.

When the responses of all the subjects are combined, a rather interesting result emerges: The total consistency (for all 27 cases considered) increases significantly. The same does not hold true for each separate level of interarrival time. The implications of these findings are discussed in the following sections.

4.4.3 Interchangability of subjects and the normal distribution model

The observed result, that aggregated responses for the data at all levels of interarrival time are more consistent than that of each individuals, supports a normal distribution model of discriminal responses based on the interchangability of subjects (see Section 3.4). In other words aggregating the responses of all subjects results in a larger sample size, and thus in a better estimate of modal discriminal responses (scale values) than the responses of each subject taken alone; and the increase in the number of trials results in a higher overall degree of consistency (see Section 4.4.1). Due to better consistency, the aggregate scale of subjective MWL given in Figure 3.4 is preferable to the individual scales of Figures 3.3.

However, the same argument does not hold for the data at short interarrival times. Here, the aggregate degree of consistency decreases. This result is interpreted in Section 4.4.5.

4.4.4 Inconsistency, disagreement, and confusion: Effects on the distances on the MWL scale

Comparing the results for each level of interarrival time it is observed that at longer interarrival times where MWL is lower, a higher degree of consistency for the aggregate data (Table 5.8) is accompanied by both a higher value for the coefficient of concordance (Table 4.7), and a larger distance between the adjacent cases on the aggregate scale of subjective MWL of Figure 3.4.

On the basis of the preceding observation, the coefficient of consistency may be thought of as an index of confusion or fuzzyness. Then it could be said that a higher concentration of stimuli in some unit length of psychological continuum is associated with higher confusion for each subject and lesser agreement among the subjects.

The diagram given below summarizes the above relationships:

Higher interarrival time →Lower MWL→ -Higher consistency
-Better agreement
-Larger distances between scale values

4.4.5 Stochastic unidimensional vs. multidimensional model of mental workload

In Section 4.4.1 two underlying models of psychological attribute which result in inconsistency of responses were discussed. The improved degree of consistency for the aggregate data of all levels of interarrival time fits the normal distribution model of discriminal responses (44 inconsistent triads and K = .94 in Table 4.9). However, for the aggregate data at the short interarrival time (high MWL) the degree of inconsistency is higher than that of each individual. There are two possible explanations for the dissimilar results at each level of interarrival time:

i) Multidimensional MWL: If subjective MWL is a multidimensional psychological attribute, then, the results of Table 4.9 indicate that there exists one dominant dimension for MWL at medium and long interarrival times (medium and low MWL regions), but all the dimensions are important for short interarrival time (high MWL region).

ii) Unidimensional MWL: If subjective MWL is a unidimensional psychological attribute, then, the results of Table 4.9 indicate that the distances between the theoretical scale values of MWL are larger than the "just noticable difference" at medium and long interarrival times (medium and low MWL region), but equal or smaller than the "just noticable difference" at short interarrival time (high MWL region).

The author believes that the second model presents a sounder interpretation of the results. The author also postulates that even if subjective MWL is a function of many independent variables (dimensions), subjects report their "weighted" judgments of MWL. Or, in other words, subjects incorporate some internal criteria to weight different dimensions of MWL and then report the total weighted subjective MWL (analogous to an internal worth function or some crude multiple regression technique), where the "weighted subjective MWL" has the properties of a unidimensional psychological attribute.

A proper statement of this postulate is that subjective mental workload, being a weighted sum of many independent variables, is associated with a unidimensional psychological continuum.

Provided that the unidimensionality of MWL is accepted, various results of the analyses suggest the possibility of an upper limit for MWL. This is discussed next.

4.4.6 Upper threshold for mental workload

At the short interarrival time (high MWL region) the results indicate a very high degree of inconsistency, very little agreement among the subjects, and an apparent lack of differentiable scale values of MWL (which results in very small distances between cases on the aggregate scale of subjective MWL). All these could be thought of as indications of an upper threshold for subjective MWL, where the spacing of scale values of stimuli on the Thurstone scale decreases below the just noticable difference. This conjecture supports the hypothesis of fixed mental capacity.

Also, based upon the previous observation it may be argued that a proper natural zero point for scales of subjective mental workload would be the upper threshold.

Chapter 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The purpose of this study was to examine mental workload in the context of a multi-task decision-making environment and to produce an interval scale of subjective mental workload. The experiments were conducted on Tulga's experimental paradigm using three subjects working on various experimental multi-task decision-making cases. The scaling method was based on Thurstone's judgment model and the law of comparative judgment, and was applied to the judgment data of subjects generated by the method of paired comparisons. Various statistical methods of analysis were employed in order to study properties both of the original data and of the scales of subjective mental workload.

The experimental environment and the results of the subjective judgments provide support for a conceptual model of subjective mental workload with the following general properties:

- subjective mental workload is stochastic in nature.
- subjective mental workload is a property of the environment - operator interface, and it depends on both environment and operator characteristics.
- subjective mental workload can be quantified on a unidimensional scale, as demonstrated by this work.

- There exists a substantial degree of agreement among subjects in their judgments of mental workload.

The following more specific conclusions can be drawn from the study:

1. Among the experimentally controlled variables, the interarrival time between tasks, T, has the most significant effect in the variation of subjective mental workload. A longer interarrival time, which is associated with a lower number of tasks-to-be processed per unit of time, results in a lower subjective mental workload.

2. In general, a lower number of tasks-to-be processed per unit time (a condition associated with longer interarrival times) is associated with the following:

- lower subjective mental workload.
- higher degree of consistency (more transitive responses) within subjects.
- better agreement among subjects on rank ordering of cases with respect to their associated mental workloads.
- larger intervals between adjacent cases on the aggregate scale of subjective mental workload.

3. Analysis of consistency indicates a high degree

degree of transitivity of responses for the aggregate data except for short interarrival times of tasks, which supports the hypothesis of the existence of a unidimensional psychological continuum associated with the attribute of subjective mental workload for medium and low levels of mental workload.

4. Absence of agreement among the subjects with respect to the ordering of experimental cases according to their mental workload, and a very low degree of consistency (low transitivity of responses) at the shortest interarrival time indicate that either,

i) mental workload is a multidimensional phychological attribute at high levels of mental workload or,

ii) the differences between scale values of mental workload for cases become less noticeable (or differentiable)at higher levels of mental workload, thus indicating the approach of an upper threshold of mental workload. The author accepts the second interpretation.

5. Analysis of consistency resulted in a higher degree of transitivity of responses for the aggregate data than that of individual subject's data, except for the short interarrival time; thus indicating that the aggre-

gate scale of subjective mental workload is a more reliable scale than the individual scales of each subject.

6. Short and medium interarrival times are associated with cases which demand some degree of monitoring and decision making, whereas the long interarrival time is associated with cases that require little monitoring and decision making and may even provide free times between the appearances of tasks. Therefore, the nature of the decision-making environment is fundamentally different for different levels of interarrival time. At different levels of interarrival time, experimentally controlled task variables and their interactions have different effects on the variation of subjective mental workload, as discussed in detail in Chapter 4.

7. The study supports the notion that subjects form internal models of the environment based on their own abilities, dispositions, and tendencies which results in their dissimilar perceptions of mental workload. In the analysis, the underlying characteristics of the subjects was inferred from their judgment patterns. The differences of opinion among subjects with respect to mental workload occurs when the subjects' personal characteristics result in adaptations of different strategies of action. It is postulated that for environments, where only one optimal

strategy of action exists, trained subjects would be in better agreement.

5.2 Recommendations

1. The methods and techniques of investigation presented in this work could be extended and/or made more complete to serve as a methodological tool of investigation of subjective mental workload in various experimental environments.

2. The technique of scaling employed in this work can be a proper method of testing or calibration of categorical or verbal scales of the Cooper-Harper type.

3. In the present experimental paradigm the "productivity" of the human operator is adjusted by the experimentor, and thus the operator is not free to work at his/her own pace. It is suggested that the experimental paradigm be enlarged or changed to include some intellectual tasks that may require self-paced problem solving or information processing on the part of the operator. Then a proper index of performance representative of the operator's effort could be correlated to measured subjective mental workload.

4. As the analyses in Chapter 4 indicates, there exist many different environmental factors, task factors,

and operator-variable interactions that significantly affect the variation in subjective mental workloads. A possible method for the study of environment-operator effects on the mental workload is some form of continuous moment to moment recording of human operator's choice behavior and underlying strategy together with dynamics of the environment. Such recording would provide the necessary "time-line analysis" that could be studied in conjunction with the subjective measurements of mental workload.

REFERENCES

1. "Proceedings of the Symposium on Man-System Interface: Advances in Workload Study", July 31 and August 1, 1978, Airline Pilots Assn., Washington, D.C.

2. Review manuscript of 1977 NATO/AGARD Meeting in Cologne, F.R. Germany, to be published.

3. Carroll, J.D. and Chang, J.J., "How to Use INDSCAL, a Computer Program for Canonical Decomposition of N-way Tables and Individual Differences in Multidimensional Scaling," Bell Telephone Laboratories, Murray Hill, N.J., 1969.

4. Chiles, W.D., Chapter 5, "Objective Methods" in Assessing Pilot Workload, Manuscript of 1977 AGARD Meeting, to be published.

5. Cooper, G.E. and Harper, R.P., "The Use of Pilot Rating in the Evaluation of Aircraft Handling Characteristics", NASA TN-D-5123, 1969.

6. Kleinman, D.L., Baron, S. and Levison, W.H., 1971. "A Control Theoretic Approach to Manned-Vehicle Systems Analysis". IEEE Trans. Automatic Control, AC-16: 824-832.

7. Knowles, W.B., "Operator Loading Tasks". Human Factors, 1963, 5, 155-161.

8. Levine, J.M., Ogden, C.D. and Eisner, E.J., "Measurement of Workload by Secondary Tasks", Washington, D.C.: Advanced Research Resources Organization, Contract No. NAS2-9637, January 1978.

9. Levison, W.H., A Model for Mental Workload in Tasks Requiring Continuous Information Processing, in N. Moray, Ed., Mental Workload, Its Theory and Measurement, Plenum Press, N.Y., 1979.

10. McDonnell, J.D., "Pilot Rating Techniques for the Estimation and Evaluation of Handling Qualities", Wright Patterson AFB, Flight Dynamics Lab AFF DL-TR-68-76., Dec. 1968.

11. McRuer, D., Graham, D., Krendel, E. and Reisener, W., Jr., 1965. Human Pilot Dynamics in Compensatory Systems-Theory Models and Experiments with Controlled Elements and Forcing Functions Variations, AFFDL-TR-65-15, Wright-Patterson AFB.

12. Moray, M., Ed.,"Mental Workload, Its Theory and Measurement", NATO Symposium on Human Workload, Plenum Press, N.Y., 1979.

13. Palmer, E., "Pilots Manual for the 4-D Area Navigation and the Autopilot Systems in the Flight Management Project Cockpit Simulators", NASA Report No. TMX-7310, October 1975.

14. Phalen, R.M., "Automatic Control Systems", Cornell University Press, Ithaca, 1977.

15. Rolfe, J.M., The Secondary Task as a Measure of Mental Load. In W.T. Singleton, J.G. Fox, and D. Whitfield (Eds.), Measurement of Man at Work. London: Taylor and Francis, 1973, 135-148.

16. Ruffell Smith, H.P., "A Simulator Study of the Interaction of Pilot Workload with Errors, Vigilance, and Decisions", NASA Technical Memorandum 78482, 1979.

17. Sheridan, T.B., "On How Often The Supervisor Should Sample", IEEE Trans. on Syst. Sci. and Cybernetics, Vol. SSC-6, No. 2, P:140-145, April 1970.

18. Sheridan, T.B. and Ferrel, W.R., "Man-Machine Systems", MIT Press, 1974.

19. Sheridan, T.B. and Johannsen, G., "Monitoring Behavior and Supervisory Control", Plenum Press, N.Y., 1976.

20. Sheridan, T.B. and Simpson, R.W., "Toward the Definition and Measurement of the Mental Workload of Transport Pilots", Draft Report to Dept. of Transportation, January 1979.

21. Sheridan, T.B. and Stassen, H.G., Definitions, Models and Measures of Mental Workload, in N. Moray, Ed., Mental Workload, Its Theory and Measurement, Plenum Press, N.Y., 1979.

22. Stevens, S.S., "Psychophydics and Social Scaling", (Monograph), General Learning Press, Morristown, N.J., 1972.

23. Torgerson, W.S., "Theory and Methods of Scaling", John Wiley and Sons, Inc., 1958.

24. Tulga, M.K., "Dynamic Decision Making in Multi-Task Supervisory Control: Comparison of an Optimal Algorithm with Human Behavior", Sc.D. Thesis, MIT, September 1978.

25. Thurndike, E.L., "Handwriting", Teach. Coll. Rec., 11, no. 2, 1960.

26. Thurstone, L.L., "A Law of Comparative Judgment", Psychol. Rev. 1927, 34, 273-286.

27. Wesson, R.B., "Planning in the World of the Air Traffic Controller", Proceed. of the 5th Int. Joint Conf. on Artif. Intell., M.I.T., Cambridge, August 1977.

28. Wierwille, W.W. and Williger, R.C., "Survey and Analysis of Operator Workload Assessment Techniques", Systemetrics, Report S-78-101, Blacksbury, VA, September 1978.

APPENDIX A: ORIGINAL FREQUENCY RESPONSE DATA

TABLE A1: Frequency Response Data of Operator 1.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1		4	1	10	4	7	11	8	0	4	4	4	6	4	4	4	6	4	4	4	4	4	4	4	4	4	4
2	4		3	4	10	9	4	6	4	7	11	4	11	4	4	4	4	4	4	4	4	4	4	4	4	4	4
3	7	5		6	6	10	9	5	0	10	4	4	4	6	10	4	4	4	8	4	4	4	4	4	4	4	4
4	2	0	2		3	2	5	5	3	4	4	4	7	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	0	6	2	5		3	5	4	6	6	4	6	4	6	7	8	0	0	4	4	4	4	4	4	4	4	4
6	1	7	2	6	9		4	4	0	4	4	0	4	4	5	4	4	7	4	4	4	4	4	4	4	8	4
7	1	0	11	3	3	4		4	4	3	8	3	4	6	4	4	4	5	4	4	4	4	4	4	4	4	4
8	4	2	3	7	0	0	4		4	4	0	4	4	4	4	10	4	4	7	4	4	4	4	4	4	4	4
9	8	0	4	5	2	4	4	4		7	6	4	6	6	4	4	6	4	4	4	8	4	4	4	4	7	4
10	0	5	2	4	6	4	5	0	1		4	5	7	5	5	4	4	4	7	4	4	4	4	4	4	7	4
11	0	1	0	4	0	0	4	4	2	4		3	6	4	4	4	4	4	7	4	4	4	4	4	4	8	4
12	0	0	0	0	2	4	5	0	0	3	5		7	7	4	4	12	7	4	4	4	4	4	7	8	4	4
13	6	1	0	1	0	0	0	0	2	1	2	5		0	4	4	4	1	5	4	4	4	4	4	4	4	4
14	0	0	2	0	2	0	2	0	2	7	4	1	4		6	4	4	13	4	4	4	7	4	4	4	11	4
15	0	0	2	0	1	3	0	4	0	3	0	4	0	2		4	4	4	4	7	4	4	4	4	4	4	4
16	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0		2	0	4	4	0	3	0	4	6	4	4
17	2	0	0	0	4	0	0	0	2	0	0	0	0	0	0	10		3	5	4	4	4	7	4	4	4	4
18	0	0	0	0	4	1	3	0	0	0	0	1	7	3	4	4	5		4	4	5	8	4	4	4	4	4
19	0	0	0	0	0	0	0	1	0	1	5	0	3	4	0	4	3	0		5	4	4	7	5	4	4	4
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3		4	5	3	4	2	5	6
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	3	0	0		3	6	4	6	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	5	0	0	0	3	5		4	1	4	8	2
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	1	0	5	5	2	4		8	4	5	5
24	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3	4	4	7	4		4	0	3
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	6	6	0	0	0		0	2
26	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	4	0	4	4	3	4	12	3	4	4		4
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	4	6	3	6	6	4	

TABLE A2: Frequency Response Data of Operator 2.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1		4	4	4	10	2	0	2	3	6	4	5	6	5	5	7	4	5	4	4	8	4	4	4	6	4	7
2	4		4	6	6	4	4	2	0	6	7	6	4	7	4	4	7	5	4	4	4	4	4	4	4	4	6
3	4	4		0	5	4	4	3	5	6	6	4	4	6	7	4	5	4	4	4	4	4	4	4	4	4	4
4	4	2	4		4	2	4	0	4	4	4	5	5	4	6	7	6	7	6	4	4	7	4	4	4	4	4
5	6	2	3	4		4	6	2	4	6	6	6	5	10	5	4	5	4	4	4	4	4	4	4	4	4	7
6	6	4	4	6	4		1	1	0	4	5	4	4	4	4	4	6	6	7	4	4	4	4	4	4	4	7
7	4	0	0	4	2	7		3	5	4	4	4	1	4	4	4	6	7	4	4	4	4	4	4	4	5	4
8	6	6	4	4	6	3	5		3	4	7	4	4	4	6	4	4	6	6	4	4	4	4	4	4	6	4
9	5	4	3	4	4	4	3	5		7	6	4	9	4	5	4	4	4	4	4	4	4	4	7	4	4	4
10	2	2	2	4	2	0	0	0	1		4	4	7	3	6	7	2	6	5	4	4	4	4	4	6	6	5
11	0	1	2	0	2	3	0	1	2	4		5	5	4	3	4	4	4	4	4	4	4	7	4	4	4	5
12	3	2	0	3	2	0	0	0	0	4	3		5	5	0	5	6	3	4	4	4	6	4	4	6	4	5
13	2	0	0	3	3	0	3	0	7	1	3	3		1	3	6	6	2	5	5	4	10	4	4	6	4	4
14	3	1	2	4	2	0	0	0	0	5	4	3	7		4	6	5	5	7	4	4	4	7	4	4	4	4
15	3	4	1	2	3	0	0	2	3	2	5	4	5	4		4	6	4	4	7	4	4	7	4	11	4	4
16	1	0	0	1	0	0	0	0	0	1	0	3	2	2	4		2	1	2	4	4	6	5	6	4	4	5
17	4	1	3	2	3	2	2	0	4	6	4	2	2	3	2	6		5	7	4	4	4	4	7	6	6	6
18	3	3	0	1	0	2	1	2	0	2	0	5	6	3	4	7	3		4	6	4	5	4	4	4	4	4
19	0	0	0	2	0	1	0	2	0	3	0	4	3	1	0	6	1	4		4	4	3	6	6	6	4	4
20	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	0	0	2	0		7	2	0	4	4	2	4
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		0	0	4	1	2	0
22	0	0	0	1	0	0	0	0	0	0	0	2	2	0	0	2	4	3	5	6	4		4	4	5	0	2
23	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	3	0	0	2	4	4	0		4	4	0	3
24	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	1	0	2	0	0	0	0		0	0	0
25	2	0	0	0	0	0	0	0	0	2	0	2	2	0	1	0	2	0	2	4	7	3	0	4		0	4
26	0	0	0	0	0	0	3	0	0	2	4	4	4	0	0	4	2	4	4	6	6	4	4	4	4		4
27	1	2	0	0	1	1	0	0	0	2	3	3	0	0	0	3	2	0	4	0	4	6	5	4	0	0	

TABLE A3: Frequency Response Data of Operator 3.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1		8	8	5	2	4	5	1	1	4	4	4	4	2	4	3	4	4	4	4	4	4	4	4	4	4	4
2	4		8	2	0	0	4	3	3	6	4	4	4	4	6	4	4	4	9	4	4	4	4	4	4	4	4
3	8	8		2	3	4	1	4	7	7	4	4	9	4	4	4	7	4	4	4	4	4	4	4	4	4	4
4	3	6	6		0	5	4	0	4	4	4	4	7	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	6	4	5	4		0	7	0	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
6	8	8	0	11	4		7	0	6	2	7	4	8	4	4	4	4	4	8	4	4	7	4	4	4	4	4
7	7	8	7	0	5	1		9	4	4	8	6	7	4	5	4	4	4	7	4	4	4	4	4	4	4	4
8	11	9	4	4	8	4	3		1	4	4	4	4	4	4	8	5	4	4	4	4	4	4	4	4	4	4
9	7	9	5	0	3	6	0	7		6	4	4	4	9	4	10	4	4	4	4	4	4	4	4	4	4	11
10	0	2	5	0	0	6	0	0	2		9	8	7	4	8	4	8	0	4	4	7	6	4	4	4	7	4
11	0	0	0	0	0	1	0	0	0	15		7	4	6	4	4	0	4	4	4	4	4	4	4	4	6	4
12	0	0	0	0	0	0	2	0	0	0	1		3	0	6	8	0	6	8	4	4	4	4	4	9	4	4
13	0	0	3	1	0	4	1	0	0	5	0	5		4	4	11	6	0	4	4	4	8	4	4	8	5	7
14	6	4	0	0	0	0	4	0	3	0	2	4	0		6	4	0	5	4	4	4	2	4	4	4	4	4
15	0	6	0	0	0	0	3	0	0	12	0	6	0	6		4	10	11	10	4	4	7	5	4	4	4	8
16	5	0	0	0	0	0	0	0	2	0	4	4	1	0	0		6	7	4	4	8	0	4	3	11	0	4
17	0	0	1	0	0	0	0	3	0	0	4	4	6	4	2	10		6	4	4	7	6	4	4	4	5	4
18	0	0	0	0	0	0	0	0	0	4	8	6	4	3	1	13	2		7	4	8	6	7	4	7	1	7
19	0	3	0	0	0	0	1	0	0	8	0	4	0	0	2	0	0	1		13	4	0	4	4	0	4	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3		9	0	4	6	0	0	0
21	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	7		1	3	4	0	0	11
22	0	0	0	0	0	1	0	0	0	2	0	0	0	2	1	4	2	2	4	4	7		2	4	7	2	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	5	10		4	0	1	4
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	2	0	0	0		0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	1	0	1	4	4	4	5	4	4		4	1
26	0	0	0	0	0	0	0	0	0	1	6	0	3	0	0	4	3	7	0	4	4	10	7	4	8		6
27	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	4	4	9	4	0	4	7	6	

TABLE A4: Aggregate of Frequency Response Data

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1		16	13	19	16	13	16	11	4	14	12	13	16	11	13	14	14	13	12	12	16	12	12	12	14	12	15
2	12		15	12	16	13	12	11	7	19	22	14	19	15	14	12	15	13	17	12	12	12	12	12	12	12	14
3	19	17		8	14	18	14	13	12	12	14	12	17	16	21	12	16	12	16	12	12	12	12	12	12	12	12
4	8	8	12		7	9	13	5	11	11	12	13	19	12	14	15	14	15	14	12	12	15	12	12	12	12	12
5	12	12	10	12		7	18	6	11	16	14	16	13	20	16	16	9	8	12	12	12	12	12	12	12	12	15
6	15	19	6	23	17		12	5	6	10	16	8	16	12	13	12	14	17	19	12	12	15	12	12	12	16	15
7	12	8	18	7	10	12		16	13	11	20	13	12	14	13	12	14	16	15	12	12	12	12	12	12	13	12
8	21	17	11	15	14	7	12		8	12	11	12	12	12	14	22	13	14	18	12	12	12	12	12	12	12	12
9	20	13	12	9	9	14	7	16		20	16	12	19	19	13	18	14	12	12	12	16	12	12	15	12	15	19
10	2	9	9	8	8	10	5	0	4		17	18	21	12	19	15	14	10	16	12	15	14	12	12	14	20	13
11	0	2	2	4	2	4	4	5	4	23		15	15	14	11	12	8	12	15	12	12	12	15	12	12	18	13
12	3	2	0	3	4	4	7	0	0	7	9		15	12	10	17	18	16	16	12	12	14	12	15	23	12	13
13	8	1	3	5	3	4	4	0	9	7	5	13		5	11	21	16	3	14	13	12	22	12	12	18	13	15
14	9	5	4	4	4	0	6	0	5	12	10	8	11		16	14	9	23	15	12	12	13	15	12	12	19	12
15	3	10	3	2	4	3	3	6	3	17	5	14	5	12		12	20	19	18	18	12	15	18	12	19	12	16
16	6	0	0	1	0	0	0	2	2	1	4	7	3	2	4		10	8	10	12	12	9	9	13	21	8	13
17	6	1	4	2	7	2	2	3	6	8	8	6	8	7	4	26		14	16	12	15	14	15	15	14	15	14
18	3	3	0	1	4	3	4	2	0	6	8	12	17	9	9	24	10		15	14	17	19	15	12	15	9	15
19	0	3	0	2	0	1	1	3	0	12	5	8	6	5	2	10	4	5		22	12	7	17	15	10	12	8
20	0	0	0	0	0	0	0	0	0	0	0	0	3	0	2	0	0	2	6		20	7	7	14	6	7	10
21	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	4	1	3	0	8		4	9	12	7	2	11
22	0	0	0	1	0	1	0	0	0	2	0	2	2	3	1	11	6	4	9	13	16		10	9	16	10	4
23	0	0	0	0	0	0	0	0	0	0	1	0	0	1	3	7	1	1	7	9	11	14		16	8	6	12
24	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	7	1	0	5	6	4	7	4		4	0	3
25	2	0	0	0	0	0	0	0	0	2	0	5	0	0	1	3	2	1	6	14	17	8	4	8		4	7
26	0	0	0	0	0	0	3	0	1	4	10	4	7	1	0	12	5	15	8	13	14	26	14	12	16		14
27	1	2	0	0	1	1	0	0	1	3	3	3	1	0	0	3	2	1	8	6	17	16	8	13	13	10	

APPENDIX B: AN EXAMPLE OF SCALING PROCEDURE

Table B1 is an example of a verbal response matrix. The verbal response matrix is transformed into the frequency response matrix F of Table B2 according to the criteria of Table 2.1 in Chapter 2. Then, matrix F is transformed into Proportion matrix P of Table B3 by dividing each element of f'_{jk} by the sum of $f'_{jk} + f'_{jk}$. From matrix P, matrix X of Table B4 is constructed whose elements are equal to the unit normal deviate corresponding to the elements P'_{jk}. Finally, by substituting the values of x'_{jk} in Equation 3.7

$$S'_k = \frac{1}{n} \sum_{j=1}^{h} X'_{jk} \quad ,$$

The scale of Figure B1 is constructed.

Cases	1	2	3	4	5	6
1		+a+a-b+a	+a	+a-a	+a+a	+ac
		-b+a	+a	-a	-ac	-b+a
		+a+a	+a	+b	+b+a	+a
2	-a-a+b		+a+a-a	+a-a	+b+a	+ac
	+6-a		+a+a	-b-a	-a	+a-a
	-a-a		+b-b	c	-a	+a
3	-a	-a-a+a		-b-b	+a+b	+a+a
	-a	-a-a		-a-a	c -a	-a
	-a	-b+b		-a	-a	+a
4	-a+a	-a+a	=b+b		+a c	-b+a
	+a	+b+a	+a+a		-a-a	-a
	-b	c	+a		-a-a	+a
5	-a-a	-b-a	-a-b	-a c		+a+a
	+a c	+a	c +a	+a+a		+a+a
	-b-a	+a	+a	-a+a		-a
6	-a c	-a c	-a-a	+b-a	-a-a	
	+b-a	-a+a	+a	+a	-a-a	
	-a	-a	-a	-a	-a	

Table B1: An Example of Verbal Response Matrix

CASES	1	2	3	4	5	6
1		22	12	7	17	15
2	6		20	7	7	14
3	0	8		4	9	12
4	9	12	16		10	9
5	7	9	11	14		16
6	5	6	4	7	4	

TABLE B2: Frequency Response Matrix F

CASES	1	2	3	4	5	6
1		0.782	1.00	.437	.708	.750
2	.214		.714	.350	.437	.700
3	0.00	.286		.200	.450	.750
4	.563	.650	.800		.417	.563
5	.242	.563	.550	.583		.800
6	.250	.300	.250	.437	.200	

Note: $P'_{jk} + P'_{jk} = 1.00$

TABLE B3: Proportion Matrix P

CASES	1	2	3	4	5	6	
1	0.00	.79	5.00	-.16	.55	.67	
2	-.79	0.00	0.57	-.39	-.16	-52	
3	-5.00	-.57	0.00	-.84	-.13	-67	
4	.16	.39	.84	0.00	-.21	.16	
5	-.55	.16	.13	.21	0.00	.84	
6	-.67	-.52	-.67	-.16	-.84	0.00	
$\sum_{j}^{n} x'_{jk} =$	-6.85	.25	5.87	-1.34	-.79	2.86	= 0.00
$\frac{1}{h} \sum_{j}^{n} x'_{jk} = S'_k$	-1.14	.04	.97	-.22	-.13	.48	= 0.00

Note 1: $x'_{jk} + x'_{jk} = o$

2: x'_{jk} is equal to the unit normal deviate corresponding to element P'_{jk}

TABLE B4: Distance Matrix X

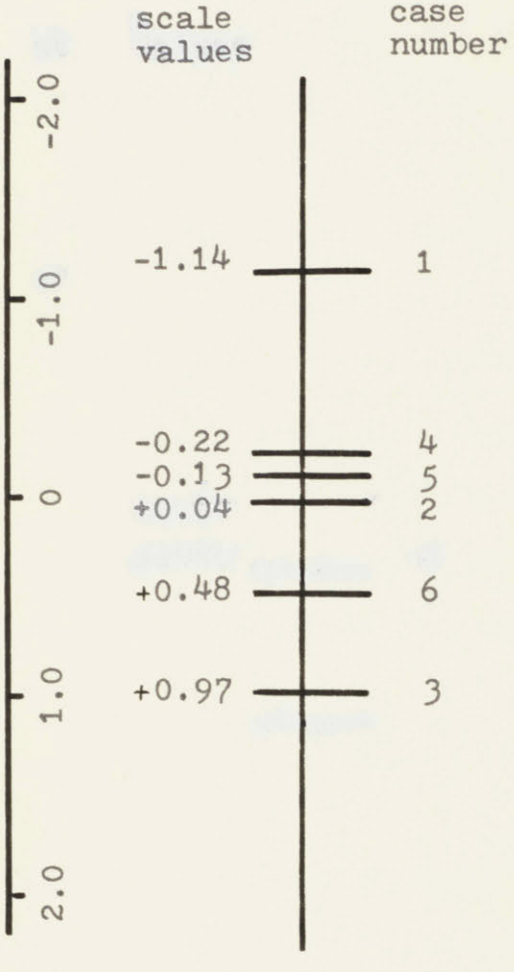


Figure B1 An Example of the Scale of Subjective Mental Workload

APPENDIX C: STATISTICAL INDICES OF AGREEMENT AND CONSISTENCY

C.1: Coefficient of Concordance

Coefficient of concordance, W,* is a measure of the agreement among subjects.

Suppose m judges rank order n stimuli from 1 to n. If the judges were unable to discriminate among the stimuli, then the expected rank total† for each stimulus would be $\frac{m(n+1)}{2}$. The difference between the observed rank totals and expected rank totals is regarded as a measure of the agreement among judges. If the judges are in perfect agreement, then the sum of the squares of the difference between observed and expected rank totals is a maximum:

$$S_{max} = \frac{m^2(n^3 - n)}{12}$$

Usually, the sum of squares, S, will be less than this amount. Coefficient of concordance, W, is defined as:

$$W = \frac{S}{S_{max}} = \frac{12\ S}{m^2(n^3-n)},$$

* For an introduction to the material of this section see: M.J. Moroney "Facts from Figures", Penquin Books Ltd.

† rank total of a stimulus is the sum of the ranks given by all judges for that stimulus.

where W can vary from 0 (complete randomness is ranking) to 1 (complete agreement).

Coefficient of concordance, W, is tested for significance using Snedecor's distribution for F, where

$$F = \frac{(m-1)\ W}{1-W} ,$$

with

Degrees of freedom for the greater estimate =

$$(n-1) - \frac{2}{m}$$

and

Degrees of freedom for the lesser estimate =

$$(m-1)\ [(n-1) - \frac{2}{m}] .$$

C.2: Spearman's Rank Correlation Coefficient

Spearman's Rank Correlation Coefficient, R, is a measure of agreement between a Pair of subjects.

Suppose two judges rank n stimuli from 1 to n. Then Spearman's Rank Correlation Coefficient is defined by:

$$R = 1 - \frac{6\sum d^2}{n^3 - n}$$

where d denotes the difference between the rankings of the two judges.

The Spearman's Rank Correlation Coefficient, R, is tested for significance using Student's t distribution where

$$\text{student's t} = R\sqrt{\frac{n-2}{1-R^2}}$$

with n-2 degrees of freedom.

C-3: Coefficient of Consistency and the Number of Inconsistant triads

Given n stimuli, judged on paired comparisons basis, a response matrix could be constructed whose elements are 0, 1, or blank. Corresponding to the elemental values of proportion matrix which are less than, greater than, or equal 0.5, respectively. In this way, the new matrix represents the average order of judgments.

The number of inconsistent triads, d, is computed in the following way:

- establish the sum of each row of the matrix
- the expected value of the sum of each row is $E = \frac{n-1}{2}$ for n stimuli. Compute $(S-E)^2$
- Compute $T = \sum_{i}^{n=rows} (S-E)^2$
- Compute $T_{max} = \frac{n^3-n}{12}$
- and finally, compute the number of inconsistent

triads, d, from:

$$d = \frac{T_{max} - T}{2}$$

Coefficient of consistency, K, is defined as:

$$K = 1 - \frac{24d}{n^3 - n} \quad \text{for n an odd number}$$

$$K = 1 - \frac{24d}{n^3 - 4n} \quad \text{for n an even number}$$

Note that a proper method of testing for the significance of K has not been developed yet.