INSTITUTE FOR THE
STUDY OF
INTELLECTUAL
BEHAVIOR

MUENZINGER BUILDING
UNIVERSITY OF COLORADO
BOULDER, COLORADO 80309

Theory Development and Experimental Evaluation in Attention Research[1)]

Franz Schmalhofer

University of Colorado

Technical Report No. 95

Institute for the Study of Intellectual Behavior

University of Colorado

February, 1980

ABSTRACT

A review of the literature reveals that the relationship between theories of attention is unclear. The present study uses a measurement theory approach in an attempt to clarify the relationship between the several different theories. Three probabilistic theories, T1, T2, and T3, each of which describes task performance as a monotone (T1), linear (T2), or logistic (T3) function of task demands and of allocated resources are proposed. Both task demands and attention are interpreted as integrals of functions in time, termed the task requirement function and the resource allocation function. A combination of the speed-accuracy trade off (SATO) and the dual task paradigm is proposed to test the theories. Unlike T2 and T3, T1 can be viewed as a formalization of Norman and Bobrow's concepts of data- and resource-limited processes. A duration discrimination experiment showed that T1 provides the best fit to the data.

Author's address: Franz Schmalhofer, Department of Psychology, University of Colorado, Boulder, Colo 80309.

Abstract

A review of the literature reveals that the relationship between theories of attention is unclear. The present study uses a measurement theory approach in an attempt to clarify the relationship between the several different theories. Three probabilistic theories, T1, T2, and T3, each of which describes task performance as a monotone (T1), linear (T2), or logistic (T3) function of task demands and of allocated resources are proposed. Both task demands and attention are interpreted as integrals of functions in time, termed the task requirement function and the resource allocation function. A combination of the speed-accuracy trade off (SATO) and the dual task paradigm is proposed to test the theories. Unlike T2 and T3, T1 can be viewed as a formalization of Norman and Bobrow's concepts of data- and resource-limited processes. A duration discrimination experiment showed that T1 provides the best fit to the data.

1) Introduction

Frequently, classes of theories about the same subject area have been developed and tested in cognitive psychology. After a controversy of which type of theory is the correct one, these classes of theories have been proven not to be distinguishable from one another on the basis of the empirical data under consideration. Incremental (Estes,1950) vs. all-or-none learning (Bower,1961), two stage (Bower,1961) vs. three stage Markovian models (Restle,1962), network (Collins and Quillian,1969) vs. set-theoretic models (Meyer,1970; Rips, Shoben and Smith 1973), serial (Sternberg,1966) vs. parallel processing (Atkinson, Holmgren and Juola,1969; Shiffrin and Schneider,1977), and imagery (Kosslyn and Pomerantz, 1977) vs. propositional memory representations (Pylyshyn,1973) were perceived as competing for the explanation of the same set of experimental data. Though these classes of theories have diametrically different features, it has been proven that they cannot be distinguished by the experimental data to which they were applied (Estes,1960; Greeno and Steiner,1964; Hollan,1975; Townsend,1974; Anderson,1979).

Similarly, there are several concepts in attention research some of which may be equivalent with respect to the experimental data. Other concepts, however, may not be comparable as they refer to different sets of experimental data. Limited central capacity, structural interference (Kahneman,1973), data- and resource-limited processes, automatization by practice (Norman and Bobrow,1975), automatic parallel and controlled serial processes (Shiffrin and Schneider,1977) and modality specific interference (Proctor,1978) are some examples of widely used concepts in attention research.

This paper attempts to clarify the relationship between these concepts of attention. First, a way of theory construction and evaluation is outlined. This method, well known but not often applied in cognitive psychology, guarantees that:

a) all predictions of a theory can be tested at once,

b) the equivalence of two theories is obvious, and

c) the psychological knowledge about one subject area is represented within one framework.

Second, a general class of probabilistic theories is stated which allows a comparison of the concepts previously developed in attention research. It is demonstrated that several concepts in attention research are empirically equivalent. Third, three specific theories are used to implicitly define attention and task demands as psychological terms. Parameter estimation and testing procedures are discussed. Finally, the proposed theories are tested by an experiment on the duration discrimination of two visual stimuli.

2) Theory Construction and Evaluation

In this section a procedure for constructing and testing theories is outlined (Block and Marschak, 1960; Hempel, 1952; Suppes and Zinnes, 1963). It will be used to compare several attention concepts and to develop three probabilistic attention theories which will implicitly define attention as a psychological concept.

Determination of an observation space: The first step in theory development is to delineate that part of the empirical world (observation space) to which the theory will apply. Thus for any given experimental design it will be possible to decide whether or not a particular theory applies. Observation spaces are chosen which are characterized by certain regularities. An observation space is given by a system (E,Q) where E is a set of experimental conditions and $Q=\{q_1,q_2, \ldots\}$ is some measure (i.e. relative frequencies) defined on E. By the definition of the observation space the semantic contents of a theory are determined.

Axiomatic theories: The observation space is related to the technical terms to be introduced by one or more axioms.

In psychology, technical terms are often related to reaction probabilities $P\{R_{a1,a2,a3,...am}\}$ and thereby to relative frequencies (Falmagne, 1978). The introduction of an axiomatic theory creates two problems (compare Suppes and Zinnes, 1963):

a) It must be decided whether or not the axioms are valid for given observations of the observation space (representation problem).

b) The empirical meaning of the technical terms must be determined (uniqueness, meaningfulness problem).

For referring to informal and less precisely developed "theories" the term "theory-class" is introduced. A theory-class is a set of possible theories which have some principle in common.

<u>Necessary and sufficient conditions for the empirical validity of a theory</u>: The following definitions distinguish clearly between identical and equivalent theories.

<u>Definition 1</u>: Two theories, T1 and T2, are <u>empirically equivalent</u> iff the sets of all possible observations O_{T1}, O_{T2}, consistent with the theories T1, T2, are identical, that is iff $O_{T1} = O_{T2}$.

<u>Definition 2</u>: A theory T1 is <u>more restrictive</u> than a theory T2 iff $O_{T1} \subset O_{T2}$ and $O_{T1} \neq O_{T2}$.

Definition 3: Two theories are identical iff their axioms are identical; they are different iff at least one of their axioms is different.

In order to test a theory T it is desirable to determine the empirical contents of T. Therefore one attempts to develop an axiomatic theory T' consisting only of observational terms which, however, is empirically equivalent to T. Craig (1953) has proven that such a substitute T' always can be developed, if the language used to talk about the empirical world is recursively enumerable. A different proof of this relation between technical terms and observables was presented by Ramsey (1931). In measurement theory the proof of the empirical equivalence between T' (representation conditions) and T is called representation theorem. Since T' contains only observational terms it can easily be tested. Furthermore, theories can be ordered by the restrictiveness of their representation conditions (Falmagne, Iverson and Marcovici, 1979). Obviously, the empirical validity of a theory implies the validity of a less restrictive theory. Concerning validity, empirically equivalent theories form an equivalence class. On empirical grounds, therefore, one cannot decide which theory is better. This decision has to be made on the basis of other considerations such as plausibility, convenience, simplicity, etc.

Uniqueness and meaningfullness: It must be decided to what degree the meaning of the introduced technical terms is determined by observational terms and axioms (uniqueness problem). Consider the two equations both known as the Bradley-Terry-Luce (BTL) axiom:

$$p(a,b) = \frac{u(a)}{u(a)+u(b)}, \qquad (1)$$

$$p(a,b) = \frac{1}{1+\exp-[v(a)-v(b)]}. \qquad (2)$$

The two equations are empirically equivalent because they can be derived mathematically from one another by substituting $v:=\log(u)$ in (1) or $u:=\exp(v)$ in (2). However, in (1) u is a rational scale, whereas in (2) v is a difference scale. Therefore, it makes sense to conceive the two axioms as different theories. In definition 4 a model is defined as the accumulated knowledge about an observation space.

Definition 4: An O-model is a 4-tuple {O,R,A,T} where O := a set of observables (observation space), R := a set of representation constraints, A := a set of axiomatic theories containing elements of the observation space, and T := a set of theorems rank ordering the axiomatic theories according to their empirical restrictions. An O-model $M_1=\{O,R_1,A_1,T_1\}$

is said to be farther developed than an O-model $M_2=\{O,R_2,A_2,T_2\}$ iff $R_2 \subset R_1$; $A_2 \subset A_1$; $T_2 \subset T_1$.

It should be noted that this kind of theory construction does not restrict the types of theories which can be developed. The condition that the language used to talk about the empirical world must be recursively enumerable only restricts one to formal descriptions. Obviously, without a formal representation no mathematical proof can be conducted. A theory construction following these rules, however, seems to be particularly useful because controversies about equivalent theories are avoided and because the theories can be tested as a whole. For any set of data it can be decided easily whether the theory is true. After briefly reviewing research on attention the procedure outlined above will be used to develop competing theory-classes of attention into axiomatic theories.

3) Experiments and Theory-Classes of Attention

Dichotic listening (Cherry, 1953; Broadbent and Gregory, 1964; Lewis, 1970), perception of temporal order (Titchener 1908; Sternberg and Knoll 1973), letter matching (Posner and Boies, 1971), dual task experiments (Keele, 1967; Schulman and Greenberg 1971) and search and scanning paradigms (Schneider and Shiffrin 1977) have been used for research on attention. There are several theory-classes to explain one or more of those experiments: Broadbent's (1958) filter model assumes strictly serial processing. Allowance for parallel processing is given by Treisman's (1960) attenuation model. Deutsch and Deutsch (1963) assume that limitations by attention are given only for response selection but not for the perceptual and discriminatory mechanisms. Posner and Boies (1971) have postulated three components of attention: alertness, selectivity and limited capacity. In visual search Shiffrin and Schneider (1977) distinguish between automatic detection and controlled search. Automatic detection can take place in parallel whereas controlled search proceeds serially. Using the notion of limited capacity, Kahneman (1973) postulates that any two tasks whose joint demands exceed that limit must be mutually interferring. Distinguishing between data- and resource-limited processes Norman and Bobrow (1975) contradict this assertion. The concept of structural interference assumes that performance may deteriorate as a result of the competition of two tasks for the same sensory or motor system (i.e. Lansman, 1978). Other authors postulate modality specific resources (Proctor,1978; Proctor and Fagnani, 1978). Since these theory-classes are in a state of development rather than being precisely formulated theories, they have to be specified more rigorously in order to identify the characteristics of attention.

On the other hand, there are mathematical theories of attention which apply only to a single paradigm (i.e. Sternberg and Knoll, 1973). Several psychologists, however, think that attention (limited capacity, consciousness, etc.) must be a basic and central concept in a cognitive theory as it plays an important role in almost all cognitive tasks (i.e. Mandler, 1975; Miller, 1956). Attention appears to be as central to cognitive activity as is the magical number seven. Therefore, a theory of attention is needed which is applicable to a wide variety of experimental tasks.

In order to compare previously developed concepts in attention research a probabilistic theory-class of attention is stated in the next section. Three specific theories of this class will be used to define attention as a psychological concept.

4) A General Class of Probabilistic Theories

The parameters of the probability distribution of all possible task performances will be used to identify both the task demands and the attention allocated to that task. First, a class of probabilistic theories, general enough to discuss the theory-classes described above within one framework, will be defined. For three particular theories of this class representation theorems will be proven.

The assumption is made that the human information processing system consists of a finite number of subsystems. Performance in a task depends on the task demands upon every subsystem and upon the attention allocated by every subsystem to that task. Assume that every subsystem i is characterized by its maximum amount of resources $\mathcal{A}_i$ available at any time instant t. This assumption is called the limited capacity property. In one form or another this assumption has been made by all attention researchers (James,1890; Kahneman,1973; Norman and Bobrow,1975; Posner and Boies,1971; Navon and Gopher, 1979). The generality of this assumption is well supported by the resluts of many different kinds of experiments (Shiffrin, 1976). Assume that resource allocation as well as task requirements are a function in time. Let $\alpha_{ij}(t)$ be a function in time describing the resources allocated by subsystem i to task j. Similarly, $\sigma_{ij}(t)$ describes the task requirements of task j from subsystem i. Let W_a and W_s be the space of all resource allocation functions and task requirement functions, respectively. For every subsystem i there exists a function H_i mapping $W_a \times W_s$ into IR, describing how well the task requirements are fulfilled by the resource allocation. A function F : IR^n --> IR describes how well the overall performance in a task depends upon the n subsystems.

This relation between task requirements, resource allocation and performance is expressed by

$$P(X_j=x_j) = G^*\{F[H_1(\alpha_{1j}(t), \sigma_{1j}(t))], \quad . \; . \; [H_n(\alpha_{nj}(t), \sigma_{nj}(t))]\}, \quad (3)$$

where G^* is a distribution function and X_j is a random variable describing all possible performances in a task j. The limited capacity assumption is expressed by

$$\alpha_{ij}(t) \leq \mathcal{A}_i, \quad \forall i,j,t; \qquad (4a)$$

$$\sum_j \alpha_{ij}(t) \leq \mathcal{A}_i, \quad \forall i,t; \qquad (4b)$$

$$\sigma_{ij}(t) \geq 0, \qquad \forall i,j,t.$$

In Eq. (4b) the index j is used to designate tasks which are performed simultaneously. Frequently, there are only two possible outcomes in the experimental task. The subject's response is either correct ($X_j=1$) or false ($X_j=0$), so that a binomial distribution function is obtained. Then Eq. (3) can be simplified by applying G^{-1}. As a consequence the left side of the equation will describe the parameter of the binomial distribution function instead of the binomial distribution function.

$$P(X_j=1) = F\{[H_1(\alpha_{1j}(t), \sigma_{1j}(t))], \ . \ . \ [H_n(\alpha_{nj}(t), \sigma_{nj}(t))]\}. \qquad (5)$$

Thus far no restrictions have been imposed upon the functions F and H, so that (5) is general enough to describe a variety of different information processing theories:

a) Assume that the subsystems 1 to n are sequentially ordered where subsystem 1 receives the physical input, operates on this input and submits its result to subsystem 2. In general, subsystem i operates on the output of subsystem i-1. Obviously, if information was lost in subsystem i this information loss cannot be recovered by additional processing of the next subsystems. This system is called a <u>sequential information processing system</u>. We would call it a <u>hierarchical system</u> if subsystems receive input from several lower ranked processors. This system is described by (3) where F is specified by:

$$F\{H_1,H_2, \ . \ . \ .H_n\} := \min \{H_i:i=1,2, \ . \ . \ n\}.$$

b) For a parallel information processing system it is assumed that overall performance is determined by the n subsystems in an independent way. The following specification of F adequately describes an <u>independent parallel information processing system</u>:

$$F\{H_1, H_2, \ldots H_n\} := \sum_{i=1}^{n} H_i .$$

c) Assume that the output of subsystem 1 is a necessary input to all other subsystems whereas parallel processing occurs in subsystem 2 to n. This system is a combination of a) and b). It is easily modelled by:

$$F\{H_1, H_2, \ldots H_n\} := \sum_{i=1}^{n} \min (H_1, H_i).$$

d) Automatic processes can be modelled by setting the task requirements $\sigma_{ij} \equiv 0 \ \forall i$ with H_i assuming its optimum value whenever $\sigma_{ij} \equiv 0$.

e) Automatization through learning as discussed by Norman and Bobrow (1975) can be accounted for by a change in the task requirement functions σ_{ij}.

These are but a few examples which demonstrate the generality of Eq.(3). The theory-class specified by (3) or (5) is not a process theory describing complex human information processing as for instance Shiffrin and Schneider's (1977) theory. Rather, it specifies how performance depends upon the "amount of processing" (mental resources) done by the human information processor. The main purpose of the present theory is to measure attention. Nevertheless, the examples given above demonstrate that theories can be developed within this theory-class which are consistent with complex process theories.

The theory-class defined by (3) was introduced in order to discuss and compare earlier theoretical developments within one framework. Before this is done some more restricted theory-classes are derived from Eq. (3). Instead of depending upon the functions α_{ij}, σ_{ij} themselves assume that performance depends only on some property of these functions, such as upon their integral over time:

$$l(\alpha_{ij},T) := \int_0^T \alpha_{ij}(t)dt, \qquad (6a)$$

$$l(\sigma_{ij},T) := \int_0^T \sigma_{ij}(t)dt. \qquad (6b)$$

$l(\alpha_{ij},T)$ is called attention. $l(\sigma_{ij},T)$ is called task demand. Thus, attention is the integral of the resource allocation function and task demand is the integral of the task requirement function. Notice that this assumption causes the two information processing modes of time sharing (serial processing) and parallel processing to become (empirically) indistinguishable. By (6) a subsystem which operates on two tasks sequentially in a time sharing mode is empirically equivalent to a parallel processing subsystem. Much effort has been spent in cognitive psychology on the empirical distinction between parallel and serial processing (compare Townsend, 1974).

Another restriction is introduced by the assumption that there is only one central processor. Substituting $G=F \circ H$, Eq. (5) is rewritten as:

$$P(X_j=1) = G[l(\alpha_j,T), l(\sigma_j,T)]. \tag{7a}$$

The limited capacity assumption is expressed as:

$$\sum_j l(\alpha_j,T) \le \mathcal{A}\cdot T, \tag{7b}$$

with j as an index for simultaneous tasks. T is not a measure of physical time but a measure of "psychological" time, which however, is monotonically increasing with physical time. In the next section several attention concepts which have been developed in cognitive psychology are interpreted within the framework of Eq. (3).

<u>Probabilistic theories applied to attention</u>: Broadbent's (1958) filter theory can be expressed by Eq. (7a) where $l(\alpha,T)$ can take on only one of two possible values, say 0 and 1, describing whether or not attention is paid to that task. The filter theory does not distinguish tasks by their demands. Thus G becomes a function in one argument. A difference to Eq. (7a), however, is that a

deterministic measure for task performance is used. The major restriction of the filter theory is

$$\sum_j l(\alpha_j,T) = 1.$$

Treisman's (1960) attenuation theory is very similar to the filter theory, the difference being that $l(\alpha_j,T)$ varies continually. As a consequence, a continuous instead of a all-or-none measure of performance is used. Kahneman's (1973) effort theory introduced task demands as a theoretical construct. Equations (7a) and (7b) together adequately formalize the effort theory. Kahneman's assertion that two tasks whose joint demands exceed the resource limits must be mutually interferring requires the function G to be strictly monotonically increasing. This assertion was refuted by Norman and Bobrow (1975) who introduced the concepts of data- and resource-limited processes. Using the present terminology these concepts reduce to the assumption that G is only monotonically nondecreasing instead of strictly monotonically (i.e. linearly) increasing. Instead of introducing a general attention-demand performance function Norman and Bobrow prefer to use a separate attention performance function F'_j (which they call performance-resource function) for every task j. Using Eq. (7a) F'_j is defined by:

$$F'_j(l(\alpha_j,T)) := G[l(\alpha_j,T),l(\sigma_j,T)].$$

If F'_j is strictly monotonically increasing within some range, performance is called resource-limited; if F'_j is a constant function it is called data-limited. Thus an important difference between Kahneman's and Norman and Bobrow's assumptions is whether to view the attention performance function as a strictly monotonically increasing or only as a monotonically nondecreasing function. Another generalization was introduced by Posner and Boies (1971) who investigated resource allocation as a time dependent process. By varying the time interval between primary and secondary task they investigated the resource allocation function α_j of a letter comparison task j (compare Eq. 7a). Because there was no interference between primary and secondary task within a given time period, they concluded that no resources were needed by the primary task at that time. In order to derive this conclusion it again must be assumed that the performance-resource function is strictly increasing. In addition, strong assumptions about the processing mode (i.e. strictly parallel processing for a given time period Δt) are necessary to make these conclusions valid. From an analysis of variance of a two factorial dual task paradigm Proctor and Fagnani (1978) concluded that there exist two subsystems, namely a visual and an auditory subsystem.

Again, in terms of Eq. (7a) this conclusion is only reasonable if G is assumed to be strictly increasing.

It is seen, that from the same and similar experimental findings, drastically different conclusions have been drawn. In many cases the assumptions leading to those conclusions are not explicitly mentioned, so that it is not clear why different researchers disagree in their conclusions about attention. Therefore, it is important to investigate the relation between different assumptions about attention.

Empirically equivalent assumptions about attention: The basic assumption is that task demands and attention must be identified by a measure of performance on these tasks. A great deal of attention research started out with the assumption that there is only one central processor. Assume that for testing this assumption a dual task paradigm was employed. Furthermore assume that the experimental results of some (n-m) tasks are adequately explained by the assumption of just one central processor, whereas the inclusion of the other m tasks invalidates this assumption. Then the following conclusions are empirically equivalent:

a) There is only one central processor. The inconsistencies in the experimental data are due to structural interference (compare Kahneman, 1973). Define the two tasks which together exert the most resources (best performances relative to task difficulty) as not being influenced by structural interference. Add all those tasks whose performance measure can be explained without structural interference. For all other tasks define the decrease in performance as structural interference with the particular task with which they are combined. By increasing the number of structures which can interfere, it is always possible to explain the experimental data in this way. Obviously, this explanation is most attractive if the number of structural interferences is very small.

b) There are several subsystems. Thus the reason why the m tasks cannot be explained by assuming only one central processor is that they either require another processor which, however, is occupied by the secondary task, or that they require an additional processor which is not occupied and thus performance is improved. Obviously, by increasing the number of information processors all experimental data will be explained. This explanation is most attractive when the dual task makes use of different stimulus modalities.

c) Assume that there is only one processor, but that resource allocation is a function in time. Then the data can be explained by defining time intervals of the resource allocation function. Task interference occurs when two tasks require resources at the same period in time.

d) Norman and Bobrow (1975) found the following two assumptions to be equivalent: Some tasks do not require any attention but the attention performance function is strictly monotonically (i.e. linear) increasing. Every task does draw from the pool of limited resources but the attention performance function is only monotonically nondecreasing.

These equivalences offer a variety of routes to develop theories of attention. Though it might be desirable to have several formally developed theories available, there is also a legitimate interest concerning which ways of developing theories are more promising. Recently, Navon and Gopher (1979) reviewed the literature on attention. They employed normative theories from economics to describe a person performing a single or dual task. Their evaluation is to a large degree consistent with the conclusions derived here by using probabilistic measurement theory. In particular, there is the agreement that "many theoretical possibilities . . . exist and that caution is needed in interpreting empirical data of dual-task performance" (p. 251). Contrary to Navon and Gopher, however, it is not the author's objective to endorse the notion of multiple resources. Because of the empirical equivalences mentioned above, the assumption of single versus multiple resources is not testable unless arbitrary assumptions are made. Navon and Gopher search for a theory which accounts for all experimental data. Therefore they need a structurally very complicated theory which, however, is too complex to be totally tested within a reasonable time period. The opposite approach of developing a simple but nevertheless reasonable theory whose validity can be determined by one or a few experiments is taken here. There is a fundamental difference between the two approaches. Instead of developing a theory which is constistent with all experimental data it is proposed to find the experimental data which are consistent with an "attractive" theory. Navon and Gopher and many others regard the experimental data as given and try to develop a theory which explains all known experimental results. On the other hand we consider the theory as given and search for the observation space for which a particular theory is valid. Once this theory is established for some domain, it can be used to investigate more complex domains. This is to say that we suggest a constructive procedure.

Thus it is proposed to design theories and experimental tasks which support the central capacity notion. Even if "in reality" multiple resources exist, one can still establish a single capacity theory by selecting experimental tasks which demand only a single or an equally weighted combination of several resources. Once a set of experimental tasks is found which supports the central capacity notion, the (established) central capacity theory can be used to identify multiple resources, structural interference, automatic processes, etc. As previously mentioned a constructive approach is taken. Rather than partially testing a complex and insufficiently specified theory-class, it is suggested to first establish a relatively simple theory for some observation space. Then the established theory can be used to investigate more complex observation spaces. In the next section three relatively simple theories of attention are developed.

5) Implicit Definition of Attention and Task Demands by Probabilistic Theories

So far we discussed attention and task demands as a function of other empirically unidentified terms, namely the resource allocation function $\alpha(t)$ and the task requirement function $\sigma(t)$. Specifically, we introduced attention $l(\alpha(t),T)$ as the integral of the resource allocation function $\alpha(t)$ and task demand $l(\sigma(t),T)$ as the integral of the task requirement function $\sigma(t)$. Thus the resource allocation function is distinguished from attention in the same way that task requirements are differentiated from task demands. In order to develop a meaningful theory of attention, task demands and attention must be identified as functions of experimental variables, or at least as functions of observables. Therefore attention and task demands will now be defined as functions of observables (i.e. experimental variables). This is done by writing $l^*(a)$ and $l^*(s)$ for attention and task demands with a and s being descriptions of the experimental variables. Attention and task demands are defined in two different ways, so that the following identities hold:

$$l^*(a) = l(\alpha, T),$$

$$l^*(s) = l(\sigma, T).$$

Notice the difference between l and l^*: l maps resource allocation functions α or task requirement functions σ into attention or task demands, whereas l^* maps experimental variables into attention or task demands. Once the distinction between l and l^* is understood, for simplicity we may use l for both functions. By relating observables to attention and task demands these concepts will be implicitly defined as psychological concepts (Hempel, 1952).

<u>Definition of Observation Spaces</u>: Assume (4a,4b) and (6a,6b) as the fundamental characteristics of the human information processor. In addition, assume that experiments can be designed so that the subject allocates all resources to the experimental tasks. Then attention can be varied experimentally by imposing different deadlines (DL) upon the subject's performance in a speed-accuracy tradeoff (SATO)

paradigm. Assume that $A_s = \{a,b,c,d, \ldots\}$ is an ordered set of DL conditions in an experiment. For example a, b, c, d might correspond to DLs of 300, 350, 400, 450 msec, respectively. Task demands can be varied by using tasks varying in difficulty. Let $S_s = \{s, t, \ldots\}$ be an apriori ordered set of tasks. Then $A_s \times S_s$ describes a set of experimental conditions in a SATO- paradigm. In every trial the subject either succeeds or fails. We say the resources dominate the task demands or vice versa. Let $Q_s := \{q(a,s): a \in A_s, s \in S_s\}$ and $P_s := \{p(a,s) : a \in A_s, s \in S_s\}$ be the relative frequencies and probabilities of success taken as the dependent variables in this experiment. Then $\langle A_s \times S_s, Q_s \rangle$ and $\langle A_s \times S_s, P_s \rangle$ are called a <u>system of single task dominance frequencies</u> and a <u>system of single task dominance probabilities</u>, respectively.

The dual task paradigm is another way of varying the attention allocated to a task. Let $S_d = \{ (s,s), (s,t),(s,u). \ldots (t,s),(t,t),(t,u) \ldots\}$ be a set of dual tasks. The DL conditions are described by $A_d = \{(a_i,\bar{a}_i),(a_j,\bar{a}_j), (b_k,\bar{b}_k) \ldots\}$. Identical indices i,j,.. indicate simultaneousness. Thus $(a_i,\bar{a}_i)$ is an experimental specification which causes the attention $l(a_i)$ and $l(\bar{a}_i)$ to be allocated to a dual task. An order is defined on A_d by: $(a_i,\bar{a}_i) \leq (b_j,\bar{b}_j)$ iff $a \leq b$. S_d is ordered by $(s,t) \leq (u,v)$ iff $s \leq v$ and $t \leq v$. Let

$P_d = \{[p(a_i,s),p(\bar{a}_i,t)], \ldots\}$, then $\langle A_d S_d, P_d\rangle$ is called a <u>system of dual task dominance probabilities</u>. Define $\langle A{*}S,P\rangle$ to be the combination of both systems $\langle A_s{*}S_s,P_s\rangle$ and $\langle A_d{*}S_d,P_d\rangle$ where $A{*}S := A_s{*}S_s \cup A_d{*}S_d$. P is the probability measure on $A{*}S$ defined by the single and dual dominance system. $\langle A{*}S,P\rangle$ is called a <u>combined dominance system</u>. Single task, dual task, and combined dominance systems are denoted by D_s, D_d, and D_c, respectively. By introducing an appropriate probability measure almost every experiment on attention can be described by a single, dual, or combined dominance system.

<u>Testable conditions for dominance systems</u>: In the sequel, the discussion will be limited to single task and combined dominance systems. Three directly testable conditions of dominance sytems, which will be shown to be empirically equivalent to three probabilistic theories of attention, are introduced.

<u>Definition 5</u>: D_s is an <u>independent limited attention system</u> iff the mapping p(a,s) from the partially ordered set $A_s{*}S_s$ to [0,1] satisfies: If $a \leq b$, $s \geq t$, then $p(a,s) \leq p(b,t)$. D_c is an <u>independent limited attention system</u> iff the mapping p from $A{*}S$ --> [0,1] in addition satisfies: $p(a_i,s) \leq p(a,s)$; if $(a_i,\bar{a}_i) \geq (b_j,\bar{b}_j)$, $(s,u) \leq (t,v)$, and $p(a_i,s) \leq p(b_j,t)$ then $p(\bar{a}_i,u) \geq p(\bar{b}_j,v)$. For brevity these order restrictions are

termed R_s^I and R_c^I.

<u>Definition</u> <u>6</u>: D_s is a <u>strong limited attention system</u> iff $\forall a,b \in A_s$ and $s,t \in S_s$:

$$p(a,s) - p(b,s) = p(a,t) - p(b,t).$$

D_c is a <u>strong limited attention system</u> iff

$$p(a_i,s)+p(\bar{a}_i,t)-p(b_j,s)-p(\bar{b}_j,t) = p(a,u)-p(b,u).$$

We will refer to these restrictions by R_s^{SO} and R_c^{SO}.

<u>Definition</u> <u>7</u>: D_s is a <u>strict limited attention system</u> iff $\forall a,b \in A_s$ and $s,t \in S_s$:

$$\frac{p(a,s)}{p(s,a)} \cdot \frac{p(s,b)}{p(b,s)} = \frac{p(a,t)}{p(t,a)} \cdot \frac{p(t,b)}{p(b,t)}.$$

D_c is a <u>strict limited attention system</u> iff

$$\frac{p(a_i,s)}{p(s,a_i)} \cdot \frac{p(\bar{a}_i,t)}{p(t,\bar{a}_i)} \cdot \frac{p(s,b_j)}{p(b_j,s)} \cdot \frac{p(t,\bar{b}_j)}{p(\bar{b}_j,t)} = \frac{p(a,u)}{p(u,a)} \cdot \frac{p(u,b)}{p(b,u)}$$

Again, the terms R_s^{SI} and R_c^{SI} will be used for brevity. Dominance systems D satisfying the restrictions R^I, R^{SO}, or R^{SI} are termed D^I, D^{SO}, or D^{SI}.

Probabilistic theories: It is the task of a theory of attention to explain the dominance probabilities in terms of attention and task demands. Three theories are introduced. Similar to the notation used to describe observation spaces subscripts s and c are used to refer to subsets of axioms of the theories T1, T2 and T3.

Theory T1 assumes task performance to be a monotone function of attention and task demand. For the single task paradigm D_s theory T1 is given by $T1_s$:

$$p(a,s) = F[l(a),l(s)], \tag{8a}$$

with F: IR^2 --> [0,1] monotonically nondecreasing in the first and monotonically nonincreasing in the second argument. Theory T1 is expanded to the combined dominance system D_c by the additional axioms:

$$\begin{aligned} p(\bar{a}_i,t) &= F[l(\bar{a}_i),l(t)], \\ p(a_i,s) &= F[l(a_i),l(s)], \\ l(a_i)+l(\bar{a}_i) &= l(a), \\ l(a),l(\bar{a}_i),l(a_i) &> 0. \end{aligned} \tag{8b}$$

Thus theory $T1_c$, the theory for the combined dominance system D_c constists of the axioms (8a) and (8b).

Theory T2 proposes that task performance is a linear function of the difference between attention and task demand. For D_s T2 consists of the axiom $T2_s$:

$$p(a,s) = F[l(a)-l(s)] = \lambda l(a)-\lambda l(s), \tag{9a}$$

with F: IR --> [0,1] being a similarity transformation. Theory T2 is applied to D_c by adding the axioms :

$$\begin{aligned} p(\bar{a}_i,t) &= F[l(\bar{a}_i)-l(t)], \\ p(a_i,s) &= F[l(a_i)-l(s)], \\ l(a_i)+l(\bar{a}_i) &= l(a). \end{aligned} \tag{9b}$$

For the single task paradigm _theory T3_ is given by axiom $T3_s$:

$$p(a,s) = \frac{l(a)}{l(a)+l(s)}. \tag{10a}$$

This theory known as BTL-model can be expanded to the dual task paradigm by the following three axioms:

$$p(\bar{a}_i,t) = \frac{l(\bar{a}_i)}{l(\bar{a}_i)+l(t)},$$

$$p(a_i,s) = \frac{l(a_i)}{l(a_i)+l(s)}, \tag{10b}$$

$$l(a_i) \cdot l(\bar{a}_i) = l(a).$$

A brief description of these theories follows. T1 is the weakest of the three theories proposed. It should be noted that this theory is a formalization of the concepts introduced by Norman and Bobrow (1975) (resource- and data-limited processes). This is not true for the theories T2 and T3. In both theories the function F is strictly monotonically increasing where F is linear or logistic. Therefore, T2 and T3 can be viewed as two different formalizations of the ideas underlying Posner and Boies' (1971) conclusions. Thus, an empirical comparison between the two different assumptions seems to be possible.

Theory T2 is closely related to the theory of linear regression. Indeed, this theory could be written by a multiple regression equation in $n = |A|+|S|$ variables:

$$E(Y(a,s)) = \beta_a y_a + \beta_b y_b + \beta_c y_c + \ . \ . \ . + \beta_s y_s + \beta_t y_t + \ . \ . \ . \ .$$

where $Y(a,s)$ is the number of correct solutions and $y_a = 1$, $y_s = -1$, $x_i = 0$ for $a \neq i \neq s$.

Theory T3 is the well known BTL- or Rasch- equation (Andersen,1973; Luce and Suppes,1965; Zermelo,1929). Having introduced the the limited capacity restriction $l(a_i) \cdot l(\bar{a}_i) = l(a)$, this theory can be applied to the dual task paradigm. Note, that this multiplicative restriction could be expressed in the form of Eq. (9b) if Eq. (2) rather than Eq. (1) were used as the basic axiom of theory T3. As mentioned earlier (2) and (1) are empirical equivalent.

Notice the following differences between T2 and T3. If p(a,s)=0.5, T3 concludes that l(a)=l(s), whereas T2 concludes that l(a)=0.5-l(s) --> l(a)$\neq$l(s). On the other hand, from p(a,s) --> 0 theory T3 derives that l(s) --> ∞ , whereas T2 requires that l(a)=l(s). Therefore apriori T2 seems to fit better for the explanation of dominance systems whereas T3 is suited for choice systems. In a choice situation equal stimulus strength results in indifference, that is p(a,s)=0.5 . However, if a subject is to master a particular task, the subject's resources l(a) must be greater than the task demands l(s). Otherwise, no meaningful solution is obtained by the subject. From these considerations, it is concluded that T2 better describes the limited attention system. On the other hand, T2 cannot be applied to choice systems because for two complementary probabilities p(a,b),p(b,a) T2 can not guarantee that p(a,b)=1-p(b,a).

Representation theorems: Thus far three theories and three directly testable conditions have been introduced for both the single task and the dual task SATO-paradigm. It will be proven that these theories are valid if and only if the respective testable conditions are true. Theorems 1 to 4 prove the empirical equivalence between the theories and the testable conditions.

Theorem 1: a) Iff the dominance system D_s is an independent limited attention system, then a function l: $A_s \times S_s$ -> IR can be introduced by $T1_s$ such that l(a) is a measure of attention and l(s) is a measure of task demands.

b) The measure of attention l(a) is unique up to monotonic transformations. Also, the measure of task demands l(s) is unique up to monotonic transformations.

Proof: If D_s satisfies $T1_s$, then $D_s = D_s^I$. Conversely, if $D_s = D_s^I$, then define l(b) > l(a) iff b > a and l(s) > l(t) iff s > t. The function F is now defined by: [l(a),l(s)] -> p(a,s). For D_s^I it is seen that F is monotonically increasing in the first and monotonically decreasing in the second argument.

QED

Theorem 2: a) For D_c^I a measure of undivided l(a) and divided attention $l'(a_i, \bar{a}_i) = [l(a_i), l(\bar{a}_i)]$ is introduced by $T1_c$.

b) The measure of task demands is unique up to monotonic transformations. The measure of attention is unique up to transformations which preserve both, the order of $x,y \in A$, for which there exist $s,t \in S$ such that $s \leq t$ and $p(x,s) < p(y,t)$ and the limited capacity equation $l(a_i)+l(\bar{a}_i)=l(a)$.

<u>Proof</u>: If D_c satisfies $T1_c$, $D_c=D_c^I$. Conversely, if $D_c=D_c^I$, define $l(a)$ and $l(s)$ as in theorem 1. The remaining $l(x)$, $l(y)$ are defined, such that if $s \leq t$ and $p(x,s) < p(y,t)$ then $l(x) > l(y)$, where x and y may represent any a,b,b_i, or $\bar{b}_i$. From this definition it becomes clear that F is monotonically increasing in the first and monotonically decreasing in the second argument. Once this partial order among the values $l(x)$, $l(y)$ is determined, they can be monotonically transformed, such that $l(a_i)+l(\bar{a}_i)=l(a)$. The uniqueness property follows from these considerations.

QED

<u>Theorem 3</u>: a) Iff D_s is a strong limited attention system, then a function $l: A \times S \to IR$ can be introduced by $T2_s$ such that $l(a)$ and $l(s)$ is a derived measure of attention and task demand. Attention and task demands are measured on one scale.

b) $l(x): x \in A \cup S$ is unique up to linear transformations, that is $\langle D_s, T2_s, l \rangle$ is a derived interval scale.

<u>Proof</u>: a) "-->": D_s is a single task dominance system satisfying $T2_s$, then $\forall a,b \in A_s\ s,t \in S_s$:

$$p(a,s)-p(b,s) = \lambda l(a)-\lambda l(s)-\lambda l(b)+\lambda l(s) = \lambda l(a)-\lambda l(b),$$

$$p(a,t)-p(b,t) = \lambda l(a)-\lambda l(t)-\lambda l(b)+\lambda l(t) = \lambda l(a)-\lambda l(b).$$

"<--": D_s is a strong limited attention system. Then for two arbitrary $o \in A_s$, $z \in S_s$ define: $l(o):=\mu$; $l(z):=\lambda l(o)-p(o,z)$ $0 \neq \lambda$, $\mu \in \mathbb{R}$. For all other $a \in A_s$, $s \in S_s$ define:

$$l(a) := \frac{1}{\lambda} p(a,z)+l(z)$$

$$l(s) := l(o)- \frac{1}{\lambda} p(o,s)$$

It is to show that these definitions are valid.

$$p(a,s) = p(a,z)-p(o,z)+p(o,s),$$

$$p(a,s) = \lambda l(a)-\lambda l(z)-\lambda l(o)+\lambda l(z)+\lambda l(o)-\lambda l(s),$$

$$p(a,s) = \lambda l(a)-\lambda l(s).$$

b) It is claimed that for every two functions l and l^+ which are properly constructed there exist numbers $\gamma \neq 0$ and δ: $\forall x \in A_s \cup S_s\ l^+(x) = \gamma \cdot l(x)+\delta$. Assume there are $a,b \in A_s$; $s \in S_s$, such that

$$l^+(a) = \gamma \cdot l(a)+\delta$$

$$l^+(s) = \gamma \cdot l(s) + \delta$$
$$l^+(b) \neq \gamma\, l(b) + \delta$$

Then there is a $\eta \neq 0$: $l^+(b) = \gamma \cdot l(b) + \delta + \eta$. By $T2_s$:

$$p(a,s) = \lambda[l(a)-l(s)] = \lambda'[l^+(a)-l^+(s)],$$
$$p(b,s) = \lambda[l(b)-l(s)] = \lambda'[l^+(b)-l^+(s)],$$
$$\lambda[l(a)-l(s)] = \lambda'[\gamma l(a) + \delta - \gamma l(b) - \delta].$$

$$\text{Hence} \quad \frac{\lambda}{\lambda' \cdot \gamma} = 1,$$

$$\lambda[l(b)-l(s)] = \lambda'[\gamma l(b) + \delta + \eta - \gamma l(s) - \delta],$$

$$1 = \frac{\lambda}{\lambda' \cdot \gamma} = \frac{l(b) + \eta/\gamma - l(s)}{l(b)-l(s)} \quad \text{-->} \quad \eta = 0.$$

QED

<u>Theorem</u> <u>4</u>: a) Iff D_c is a strong limited attention system then a function $l: A_s \cup S_s \to IR$ and $l': A_d \cup S_d \to IR^2$ where $l'(x,y) := (l(x), l(y))$ can be introduced by $T2_c$ such that $l(a)$, $l(a_i)$, $l(\bar{a}_i)$, and $l(s)$ are derived measures of attention and task demand.

b) $l(a)$, $l(s)$ is unique up to linear transformations.

<u>Proof</u>: a) "-->": By insertion of T2 into the representation condition it may be seen that any dominance system satisfying T2 is a strong limited attention system.

"<--": If D_c is a strong attention system, $\forall (a,s) \in A_s \times S_s$ define $l(a)$ and $l(s)$ as in theorem 2. For any $(a_i, \bar{a}_i) \in A_d$ it is defined:

$$l(\bar{a}_i)=\frac{1}{\lambda}\ p(\bar{a}_i,z)+l(z),$$

$$l(a_i)=\frac{1}{\lambda}\ p(a_i,s)+l(s).$$

These definitions have to satisfy T2. By definition 6, $\forall(a_i,\bar{a}_i),(b_j,\bar{b}_j)\in A_d$, $a\in A_s$, $s\in S_s$ and $(s,t)\in S_d$:

$$\begin{aligned} p(\bar{a}_i,t)&=p(a,r)-p(b,r)-p(a_i,s)+p(b_j,s)+p(\bar{b}_j,t)\\ &=\lambda l(a)-\lambda l(a_i)-\lambda l(t)-\lambda l(b)+\lambda l(b_j)+\lambda l(b_j)\\ &=\lambda l(\bar{a}_i)-\lambda l(t). \end{aligned}$$

b) analogous to theorem 3.

QED

__Theorem__ __5__: a) Iff D_s (D_c) is a strict limited attention system then a function l: $A_s \cup S_s \to IR$ and a function l': $A_d \cup S_d \to IR^2$, where $l'(x,y) := (l(x),l(y))$ can be introduced by $T3_s$, $(T3_c)$ such that $l(a)$, $l(a_i)$, $l(\bar{a}_i)$, and $l(s)$ are derived measures of atttention and task demand.

b) $l(a)$ and $l(s)$ are unique up to similarity transformations. $\langle D_s,T2_s,l\rangle$ and $\langle D_c,T2_c,l\rangle$ are derived ratio scales.

__Proof__: For slightly different conditions this theorem has been proven by Luce and Suppes (1965) and Suppes and Zinnes (1963). The proof proceeds in an analogous fashion to that of the proofs of theorems 3 and 4. The initial definitions are replaced by:

$$l(o) := \mu, \quad \text{with } \mu > 0$$

$$l(z) := \frac{p(z,o)}{p(o,z)} \cdot l(o)$$

$$l(a) := \frac{p(a,z)}{p(z,a)} \cdot l(z)$$

$$l(s) := \frac{p(s,o)}{p(o,s)} \cdot l(o)$$

For the dual task dominance system we define:

$$l(a_i) := \frac{p(a_i,z)}{p(z,a_i)} \cdot l(z)$$

$$l(\bar{a}_i) := \frac{p(\bar{a}_i,s)}{p(s,\bar{a}_i)} \cdot l(s)$$

QED

Both Norman and Bobrow (1975) and Navon and Gopher (1979) proposed to use different pay-off schemes to induce changes in the subject's resource allocation. Apriori, it is not clear whether humans have voluntary control over their resource allocation. The following corollary provides a method for testing this assumption.

<u>Corollary</u>: Assume $T3_c$ is true. Then it can be tested whether a subject is able to allocate a given fraction x of his attention to the primary task and the remainder 1-x to the secondary task.

<u>Proof</u>: $l(a_i) := x \cdot l(a)$, $l(\bar{a}_i) := (1-x) \cdot l(a)$, with $0<x<1$. Then

$$p(a_i,s) = \frac{x \cdot l(a)}{x \cdot l(a)+l(s)}$$

$$\frac{p(a_i,s)}{p(s,a_i)} \frac{p(t,\bar{a}_i)}{p(\bar{a}_i,t)} = \frac{x \cdot l(a) \cdot l(t)}{l(s) \cdot (1-x) \cdot l(a)}$$

QED.

Having proven representation theorems for each of the three theories it is now possible to order these theories according to their empirical restrictions.

<u>Theorem 6</u>: For apriori unordered sets A, S the following relations hold: a) T2 and T3 are empirically more restrictive than T1 . b) T2 and T3 are not ordered by

empirical restrictions.

<u>Proof</u>: Consider the theories explaining single task experiments first.

a) Obviously, $O_{T1s} \neq O_{T2s}$ and $O_{T1s} \neq O_{T3s}$. It remains to be shown that for D_s^{SO} and D_s^{SI} there exists an order on A and S such that (a,s) --> p(a,s) is increasing in the first and decreasing in the second argument. If $T2_s$ is valid, order the $a \in A_s$ according to the order of p(a,s). Since p(a,s) = p(a,t)+p(b,s)-p(b,t), this order is independent of s. The same argument holds for the ordering of S_s. If $T3_s$ is valid, order the elements of A_s and S_s by p(a,s)/p(s,a). Since p(a,s) --> p(a,s)/p(s,a) is strictly monotone, the mapping (a,s) --> p(a,s) is increasing in the first and decreasing in the second argument.

b) is shown by a numerical example. p(a,s)=0.6, p(a,t)=0.7, p(b,s)=0.4, and p(b,t)=0.5 satisfy $T2_s$, but not $T3_s$. On the other hand p(a,s)=0.9, p(a,t)=0.75, p(b,s)=0.5 and p(b,t)=0.25 satisfy $T3_s$, but not $T2_s$.
By the same arguments, these relations hold up for combined dominance systems.

QED.

Even for an independent, strong, or strict limited attention system the relative frequencies collected in an experiment may not satisfy the restrictions R^I, R^{SO}, or R^{SI}. A statistical decision is therefore needed to determine whether the collected data originate from dominance systems satisfying these restrictions.

6) Estimation and Testing

Before parameter estimations and statistical tests are described separately for every theory, the rationale of the decision rules which are applied to all three theories is stated. Note, that the restrictions R^I, R^{SO} and R^{SI} were proven to be necessary and sufficient conditions for the correctness of the theories T1, T2 and T3 respectively. In order to test whether the relative frequencies originate from probabilities satisfying those restrictions an unconstrained estimate and an estimate satisfying the restrictions of the theory are calculated. That subset of the parameterspace Ω for which the restrictions of the theory are satisfied is referred to by ω. In the event that the theory is correct, both estimates should be statistically identical. Thus in order to test the proposed theories, a statistical test will be employed which determines whether the two estimates are identical in the statistical sense. Since the unconstrained estimate is directly given by the relative solution frequencies, only the constrained estimate need to be described. For theories T1 and T2 maximum likelihood (ML) and least square estimations are proposed. A regular best asymtotic normal (RBAN) estimation is used for T3. The RBAN estimation has a smaller sampling variance and a smaller sampling error than the ML estimate (Berkson, 1953). Also, it has the same asymtotic properties as the ML estimate (Neyman, 1949). Most important, however, for the RBAN estimation an algorithm is known which calculates the estimation values. The RBAN estimation can be expressed as a quadratic programming problem. For the ML estimation of T3 a computer program which searches for the best parameters has been used (i.e. Scheiblechner, 1979). When run on a computer the estimations usually stabilize after several iterations. The estimation is not convergent in the mathematical sense, however. It does not satisfy:

$$\forall \varepsilon \ \exists N : |p-p_n| \leq \varepsilon \ \forall n \geq N$$

For these ML algorithms there is no guarantee that the values maximizing the objective function are found by the algorithm. The superiority of the RBAN over the ML estimation is stressed because this estimation procedure has been ignored in the psychological literature. A detailed technical discussion can be found in Berkson (1944,1953) and

Taylor (1953).

Algorithms and decision characteristics for the three theories are discussed next. Estimation procedures for the dual task paradigm will not be treated until all single task estimations are established. Then it will be seen that the proposed methods can be generalized to the dual task paradigm.

6.1) Theory T1

First, a description of the minimum lower sets algorithm which is used to estimate the probabilities p(a,s) of an independent limited attention system D_s^I is given. Then several characteristics of this estimation, leading to the proof that this algorithm furnishes least square and ML estimates constrained by the order restrictions R_s^I of the system D_s is given. An extensive treatment of estimations and tests of order restrictions has been presented by Barlow, Bartholomew, Bremner and Brunk (1972). To facilitate the description of the minimum lower sets algorithm a few definitions are needed.

Definition 8: A subset $L \subset A \times S$ is a lower set with respect to the partial order defined by R_s^I iff:

$$\forall (a,s) \in L,\ b \in A,\ b \leq a \dashrightarrow (b,s) \in L,$$

$$\forall (a,s)\in L,\ t\in S,\ t\geq s \rightarrow (a,t)\in L.$$

$\mathcal{L}$ designates the set of all lower sets of A*S. A subset $U\subset A{*}S$ is an <u>upper set</u> iff $\exists\ L\subset A{*}S : U=A{*}S\cap L^c$. $\mathcal{U}$ designates the set of all upper sets of A*S. $B\subset A{*}S$ is a <u>level set</u> iff $\exists$ L,U: $B=L\cap U$. $\mathcal{B}$ stands for the set of all level sets.

Notice, that A*S and {} are lower as well as upper and level sets. In addition, all upper and lower sets are level sets. $\mathcal{L}$, $\mathcal{U}$ and $\mathcal{B}$ are closed relative to union and intersection. Now the minimum lower set algorithm can be described easily. The minimum lower set algorithm obtains estimates $\hat{p}(a,s)$ by selecting the level set B with the minimum average solution frequency from a set $\mathcal{B}'$ consisting of appropriately chosen sets of level sets.

<u>Minimum Lower Sets Algorithm</u>

1) Declare $\mathcal{L}$ as the active $\mathcal{B}'$.
2) For every set $B\in\mathcal{B}'$ calculate the relative solution frequency $\bar{q}(B)$.
3) Select the set $\hat{B}$ such that $\bar{q}(\hat{B})=\min\{\bar{q}(B)\}$. In case of indeterminancy take the union of all $\hat{B}$: $\bar{q}(B)=\min$.
4) The set $\hat{B}$ determined in 3) is called the i-th level set B_i, where i is the number of loops of this algorithm

previously executed. $\forall (a,s) \in \hat{B}$, $\bar{q}(B)$ is taken as estimate of p(a,s). Thus $\hat{p}(a,s)=\bar{q}(B)$ for $(a,s) \in \hat{B}$.

5) Declare $\mathcal{B}' = \{(\bigcup B_i)^c \cap L: B_i$ previously selected, $L \in \mathcal{L}\}$ as the active set.

6) Loop through steps 2 to 5 until $\mathcal{B}'$ contains only the empty set {}. Then an estimate $\hat{p}(a,s)$ has been assigned to every $(a,s) \in A \times S$.

Note that all estimates $\hat{p}(a,s)$ are obtained by averaging weighted relative frequencies. Thus $0 \leq \hat{p}(a,s) \leq 1$. Also, by step 3 and 5 it is guaranteed that the restrictions R_s^I are satisfied. The estimates are unique. The estimation obtained by the minimum lower sets algorithm is called the <u>isotonic regression</u> of $\bar{q}$ with weight function n and order restrictions R_s^I. In the sequel, functions satisfying R_s^I will be called <u>isotonic functions</u>. The next step is to prove that this isotonic regression supplies the least square estimate.

<u>Lemma 1</u>: a) A least square estimation subject to partial order restrictions (for example R_s^I) exists. In particular, it is unique. b) The least square estimates are obtained by averaging over appropriately selected subsets of the empirical data. In other words, if the order restricted least square estimate is a constant for some set B, this constant is the average over the relative solution

frequencies of B.

<u>Proof</u>: a) To verify (a) represent every empirical datum by one dimension in a n-dimensional space. Then the order restictions R_S^I define a convex set V in this space and the empirical data are represented by a single n-dimensional vector v. Since there exists a minimum distance line from v to V, the least square estimation is unique.

b) To prove the second part, assume that $\forall$ $(a,s) \in B \subset A \times S$ the least square estimate $\hat{p}$ takes on the value d and

$$\overset{*}{p}(a,s) = \frac{\sum_{(a,s)\in B} \bar{q}(a,s)n(a,s)}{\sum_{(a,s)\in B} n(a,s)} \neq d$$

The objective function is given by

$$\sum [\bar{q}(a,s)-\hat{p}(a,s)]^2 n(a,s)^2 =$$

$$\sum_{(a,s)\notin B} [\bar{q}(a,s)-\hat{p}(a,s)]^2 n(a,s)^2 + \sum_{(a,s)\in B} [\bar{q}(a,s)-d]^2 n(a,s)^2 .$$

Since the last term as a function of d is minimized by $d=\overset{*}{p}(a,s)$, an isotonic function coinciding with $\hat{p}(a,s)$ for $(a,s) \notin B$, and having a smaller objective value for $(a,s)\in B$ can be found. This is a contradiction to the above assumption.

QED.

Theorem 7: The least square estimates of the dominance probabilities p(a,s), constrained by the order restrictions R_S^I, are supplied by the minimum lower sets algorithm.

Proof: $[B:\hat{p}=x]$ and $[B:\hat{p}\leq x]$ define subsets $B\subset A\times S$ for which the constrained least square estimation assigns the value x or a value smaller or equal to x, respectively. The uniqueness of this definition is quaranteed by lemma 1a. $AV[B:\hat{p}=x]$ describes the average solution frequency of the set B. By lemma 1b: $AV[B:\hat{p}=x] = x$. Let x, y be two values of $\hat{p}$ such that $\hat{p}=x \ \forall (a,s)\in B_1$, $\hat{p}=y \ \forall (a,s)\in B_2$ and $[B:\hat{p}=z] = \{\}$ $\forall x<z<y$, then x and y are called adjecent values of the function $\hat{p}$. It is

$$AV[B:\hat{p}=y] = AV([B:\hat{p}\leq y]\cap[B:\hat{p}>x])$$

$$= AV(L\cap[B:p>x])$$

for all L satisfying $L\cap[B:\hat{p}>x]\neq\{\}$. This follows immediately from the definition of lower sets, and the properties of the averaging operation: $[B:\hat{p}=y]\subset(L\cap[B:\hat{p}>x])\neq\{\}$. x<y are adjecent values. For the same reason, $AV[B:\hat{p}=y]<AV(L\cap[B:\hat{p}>x])$ for all $L\in\mathfrak{L}$, satisfying $L \supset [b:\hat{p} \leq y]$ and $L \not\supset [B:\hat{p} \leq y]$ Thus $[B : \hat{p}=y]$ is the largest level set of the form $L \cap [B : \hat{p}>x]$, $L \in \mathcal{L}$ which has the minimum average of all level sets as described in the minimum lower sets algorithm. This establishes the induction step. It is also clear, that the first set is correctly selected by the minimum lower sets algorithm.

QED.

Theorem 7 has proven that the isotonic regression provides least square estimates for an independent limited attention system. The following three lemmas will be helpful in proving that the ML estimate coincides with the least square estimate.

<u>Lemma 2</u>: The following relations hold between the isotonic regression $\hat{p}$ and any other isotonic function f.

$$(\bar{q}-\hat{p})(\hat{p}-f)n^2 \geq 0 \qquad (11)$$

$$(\bar{q}-\hat{p})\hat{p}\cdot n^2 = 0 \qquad (12)$$

$$(\bar{q}-\hat{p})\cdot f\cdot n^2 \leq 0 \qquad (13)$$

<u>Proof</u>: $f' := (1-\alpha)\hat{p}+\alpha f$ is an isotonic function for $0 \leq \alpha \leq 1$, because the class of isotonic functions forms a convex set.

The sum of squared deviations of f' from $\bar{q}$ is

$$\sum \{\bar{q}-[(1-\alpha)\hat{p}+\alpha f]\}^2 n^2_{.}$$

Consider this term as a function in α . It is minimized by $\alpha=0$, because $\hat{p}$ is defined as the least square solution. Therefore the derivative of f' at $\alpha=0$ must be nonnegative. Equation (12) is seen to be true by the substitution $f=c\cdot\hat{p}$ with $c>1$ and $c<1$ in (11). Finally (13) follows from (11) and (12).

QED.

<u>Lemma 3</u>: For every real valued function Ψ

$$\sum (\bar{q}-\hat{p})\Psi(\hat{p})\cdot n = 0.$$

<u>Proof</u>: By lemma 1b $\sum_{\hat{p}=c} (\bar{q}-\hat{p})\cdot n = 0.$

QED.

<u>Lemma 4</u>: The discrepancy measure

$$\sum_{(a,s)\in A\times S} \Delta\,[\bar{q}(a,s),f(a,s)]\cdot n(a,s)$$

with

$$\Delta(\bar{q},f) = \Phi(\bar{q})-\Phi(f)-(\bar{q}-f)\cdot\varphi(f)$$

obtains its minimum value for the maximum likelihood estimate $f=\hat{p}$. The set of admissable functions f forms a closed set. φ is the derivative of Φ, which is defined by $\Phi(u)=u\cdot\log(u)+(1-u)\log(1-u)$.

Proof: $\varphi(u)=\log(u)-\log(1-u)$. By insertion,

$$\Delta(\bar{q},f) = \Phi(\bar{q})-f\cdot\log(f)-(1-f)\cdot\log(1-f)-(\bar{q}-f)\cdot[\log(f)-\log(1-f)]$$

$$= -\bar{q}\cdot\log(f)-(1-\bar{q})\cdot\log(1-f)+\text{con}.$$

Therefore, $\Delta(\bar{q},f)$ is minimized by the same function f which maximizes $\bar{q}\cdot\log(f)+(1-\bar{q})\cdot\log(1-f)$.

QED.

Theorem _8_: The isotonic regression furnishes the ML-estimation for dominance systems constrained by the restrictions R_S^I of an independent limited attention system.

Proof: By the previous lemma it is sufficient to show that Eq.(14) holds for all isotonic functions f with $0\leq f\leq 1$.

$$\sum\Delta(\bar{q},f) = \Delta(\bar{q},p)+\Delta(p,f) \tag{14}$$

By insertion it is obtained that

$\Delta(\hat{p},f) = (\bar{q}-\hat{p})\,[\varphi(\hat{p})-\varphi(f)]$. Observe, that φ is nondecreasing. Therefore, $\varphi(f)$ is isotonic. It is

$$\sum (\bar{q}-\hat{p})\cdot\varphi(f)\cdot n \leq 0 \qquad \text{by lemma 2a,}$$

$$\sum (\bar{q}-\hat{p})\cdot\varphi(\hat{p})\cdot n = 0 \qquad \text{by lemma 3.}$$

Thus $\sum \Delta(\bar{q},f)>0$ for every function f except for $f=\hat{p}$.

QED.

The last theorem establishes the constrained ML-estimation for theory T1. Hence, the likelihood ratio (LR) coefficient can be calculated by:

$$\lambda = \frac{Li\{\bar{q}(a,s)\cdot n(a,s);\bar{q}(b,s)\cdot n(a,s)\ldots \mid \hat{p}(a,s);\hat{p}(b,s)\ \ldots\}}{Li\{\bar{q}(a,s)\cdot n(a,s);\bar{q}(b,s)\cdot n(a,s)\ldots \mid p^{*}(a,s);p^{*}(b,s)\ \ldots\}}$$

with $\hat{p}\in\omega$ being the parameter space constrained by R_s^I and $p^{*}\in\Omega$ the total parameter space. Unless the empirical data $\bar{q}$ satisfy the order restrictions R_s^I $\hat{p}$ will always lie on a boundary of ω. For this case Chernoff (1954) has shown that the statistic $-2\log\lambda$ can be used to test whether the true p lies in ω (T1 is true), or $p\in\omega^c$(T1 is invalid). Under the assumptions that p lies on a boundary surface Chernoff has proven that $-2\log\lambda$ is χ^2 distibuted with 1 degree of freedom. Hereby T1 can be tested in total. Statistical procedures for T2 will be discussed next.

6.2) Theory T2

As mentioned earlier, the algorithm for estimating the β-weights in a linear regression equation could be used to estimate the probabilities p(a,s) under the side constraints of the theory T2. Indeed, it was proven by Bradley (1973) that these estimates satisfy the ML criterium. In the particular estimation problem at hand precautions would be necessary, however. No predicted value must be greater than 1. The predicted value of the null vector has to be zero. Therefore, this estimation procedure will not be pursued any further. Instead, a least square estimation is used. The objective function

$$\sum_{(a,s)\in A\times S} [\bar{q}(a,s)-\hat{p}(a,s)]^2 n(a,s)^2$$

is minimized under the side constraints $0 \leq p(a,s) \leq 1$ and R_S^{SO}. This is a quadratic programming problem. From the theory of quadratic programming it is known that except for degenerate problems there always exists a unique solution to this problem. Also, algorithms and programs are available for its solution (Land and Powell, 1973).

For testing T2, a result due to Wilks (1938) can be used. Wilks has shown that the test statistic $-2 \log \lambda$ is χ^2 distributed with k-1 degrees of freedom, if the true parameter lies in a 1 dimensional hyperplane ω of the k dimensional space Ω . λ is the LR coefficient. Obviously, $k=m \cdot n=|A| \cdot |S|$. 1 is equal to n+m-1 because any n+m-1 parameters will satisfy the restrictions R_s^{SO}. By R_s^{SO} all other probabilities are a function of the above parameters. Thus the dimensionality of ω is n+m-1. Also, Pearson's χ^2 test could be applied to test this theory. The degrees of freedom for this latter test are $(n-1)\cdot(m-1)$ since n+m-1 independent parameters have been estimated from $n \cdot m$ data. Thus both test statistics are χ^2 distributed and have the same number of degrees of freedom. As a matter of fact, the two tests have been shown to be equivalent in the limit (Neyman, 1949).

6.3 Theory T3

As mentioned earlier we will use an RBAN estimation which has been described by Berkson (1944,1949). This estimation requires the minimization of the term:

$$\sum_{a \in A} \sum_{s \in S} n(a,s)\bar{q}(a,s)[1-\bar{q}(a,s)] \left(\log \frac{\bar{q}(a,s)}{1-\bar{q}(a,s)} - \log \frac{\hat{p}(a,s)}{1-\hat{p}(a,s)} \right)^2$$

under the side constraints R_s^{SI}. The objective function as well as the side constraints R_s^{SI} can be simplified by transforming the probabilities into logits:

$$\hat{d}(a,s) = \log \frac{\hat{p}(a,s)}{1-\hat{p}(a,s)}.$$

The constraints of the optimization are now given by $(n-1)\cdot(k-1)$ linear equations:

$$\forall\, a,b,s,t, \in A,S : \hat{d}(a,s)-\hat{d}(b,s)-\hat{d}(a,t)+\hat{d}(b,t)=0.$$

Since the objective function is quadratic in $\hat{d}(a,s)$ this again can be expressed as a quadratic programming problem. Exactly the same tests as for T2 can be employed. Since the number of estimated parameters is the same, the number of degrees of freedom is identical. Hereby the discussion of statistical procedures for the single task paradigm is complete.

We will now examine whether the proposed methods generalize to the dual task paradigm . For T3 this is obvious because the objective function is still a quadratic term and the constraints are linear. Thus the estimation procedure remains a quadratic programming problem. It is not difficult to verify that the methods proposed for T2 can be generalized to the dual task paradigm as well. The order restrictions R_d^I of the theory T1 for the dual task paradigm can be expressed by a logical disjunction of several sets of partial orderings. Therefore, the istotonic regression can be applied to every factor of this logical disjunction. Thus a "ML estimate" is obtained for every factor. The result with the largest likelihood value is the ML estimate for the combined limited attention system. These statistical procedures can now be applied to determine the validity of the proposed attention theories.

7) Experimental Study[2]

A SATO single task experiment was used as first test for the proposed theories. A duration discrimination task was employed. The difficulty of this task can be varied by changing the difference between the duration of 2 visual stimuli which must be discriminated by the subject (Allan, Kristofferson and Wiens, 1971). In the present experiment this duration difference was determined by the stimulus asynchrony offset (SAO).

7.1) Method

Subjects: Four female and two male psychology students from Regensburg University (age 19-24) served as subjects.

Apparatus: The subject sat at a desk facing a vertical panel with six signal lights consisting of light emitting diodes (LED). Two signals (4 and 6) served as stimuli. The other signals were used to present feedback. Signal 1 indicated the feedback interval. Signal 2 informed the subject that his response time was longer than the lower time bound (DL_l), and signal 5 informed him that the response time was below the upper time limit (DL_u). The correctness of the response was reported by signal 3. The subject started a sequence of trials by pressing the start button. By pressing one of the two response buttons the subject decided whether signal 4 or signal 6 was lit for a longer time period. The experiment was controlled by an electronic device which was assembled to run this experiment.

Design: Five subjects were employed in a two factorial design (DL×SAO). The relative frequency of correct responses was recorded as the dependent measure. As a control, two subjects were run for ten blocks of trials with constant DL and SAO specifications.

Procedure: In order to avoid learning effects, every subject was pretrained by a few thousand trials. The pretraining was used to adjust individually the DL and SAO specifications and to familiarize the subjects with the task objectives. The subjects were instructed to use the time they had available (determined by DL_u) to make the best decision and that it was most important that their response time lie between DL_l and DL_u. In all cases DL_l was 200 msec less than DL_u. The actual experiment was run on two consecutive days.

Every trial starts with an intertrial interval (ITI) in which no signal appears. A warning interval (WI) begins when signals 4 and 6 are flashed. With equal probability the random number generator first either turns off stimulus 4 or stimulus 6. This determines the beginning of the SAO interval, which is terminated when the second stimulus disappears. It is then the subject's task to decide which stimulus lasted longer. Following the subject's response a feedback interval (FI) begins which is followed by another ITI. The subject always receives feedback regarding whether the last response was within the time bounds and whether it was correct. A sample trial sequence is shown in Figure 1.

---

insert Figure 1 about here

Between 50 and 100 trials were combined to a block. Each block contained 20 warm-up trials which were excluded from any analysis. Experimental specifications were not changed within any given block. The representation sequence of blocks was either random or ordered by the DL specifications.

7.2) Results

The data were analyzed separately for each subject because major individual differences have been found in the training phase of the experiment. More than 90 percent of all responses of a block were within the specified DL intervals. Since the remaining 10 percent were close to the DL boundaries, a well shaped unimodal distribution of latencies was obtained. The data of the two control sujects were used to test whether there were any sequential dependencies among the trials of one block. The two control subjects were presented with 10 identical blocks, each consisting of 150 trials. A time series analysis (Drosler, 1978; Yaglom,1962) showed that there were neither periodic changes (i.e. attention fluctuation) nor trends (i.e. learning or fatigue) within a sequence of 150 trials. It was also tested whether the different solution frequencies of ten identical blocks (control) can be explained by a single solution probability.

---

insert Table 1 about here

For both control and experimental subjects the results of the statistical test for a <u>single</u> solution probability are shown in table 1. The test results of the three theories are shown in Table 2. As mentioned above, the three theories have been tested for the five experimental subjects individually. The significance level was chosen to be 0.10.

insert Table 2 about here

7.3) Discussion

The solution frequencies obtained for the experimental subjects cannot be explained by a single probability. From Table 1 it can be concluded that the experimental variables (DL and SAO specifications) systematically influence the probability of a correct solution. Since no significant variation was found in the data of the two control subjects (1a and 2), it is concluded that all the systematic variability occurs in connection with the experimental variables. Therefore it must be possible to express the solution probability as a function of SAO and DL.

Three possible functional relations between SAO, DL and the solution probability were introduced by the theories T1, T2, and T3. From the results in Table 2 it is seen that the theory T1 is the most appropriate explanation. We accept T1 as an adequate explanation of the experimental results. Even though the experimental data deviate significantly from T2 and T3, these theories still explain a high percentage of the variation in the data. For instance, in the case of subject 1b, T2 and T3 explain 94.47 % and 92.62 % (determination coefficient) of the variance in the data. Therefore these theories should not be completely ruled out. Instead, some thought should be given to how error variables can be eliminated in this experiment: For subjects 4 and 5 the difference between consecutive DL specifications was only 50 msec. This had the undesired result that for DL_u=350 and DL_u=400 the average latencies were identical. Thus, the effect of the difference between the two experimental specifications vanished. Because the presentation order of experimental blocks was random for subjects 4 and 5, the solution probabilities might have been influenced by the DL specifications of the block most previously presented. Since a very powerful statistical test was applied, these error variables might have caused a type 2 error in our statistical decision about T2 and T3.

Summarizing, it is concluded that T1 adequately explains the data. Since T1 can be thought of as a formalization of "data- and resource- limited processes", these results stress the significance of Norman and Bobrow's (1975) reasoning. Though the the LR coefficient decides against T2 and T3, these theories should still be entertained as possible attention theories because the significant results could be due to error variation introduced by the random presentation order of experimental blocks, and by some experimental specifications being too similar to oneanother.

8) Conclusions

Both the theoretical analysis and the experimental study demonstrate the fact that experimental conclusions are always dependent upon the underlying mathematical theory. In attention research general conclusions have been drawn which are not true if any other than the linear model (i.e. the assumptions underlying the analysis of variance) is assumed. By restricting oneself to the linear model, many hypotheses have been rejected which may be correct when a different mathematical theory is assumed. Thereby more and more complex concepts of attention have been developed. The corresponding theory-classes became so complex, that they cannot be adequately tested within any reasonable period of time. The present paper has shown that with a different approach it is possible to develop and establish relatively simple theories of attention.

REFERENCES

ALLAN, KRISTOFFERSON & WIENS (1971) Duration discrimination of brief light flashes. *Perception and Psychophysics*. 1971,9,327-334

ANDERSEN, E.B. A goodness of fit test for the Rasch model. *Psychometrika*. 1973, 38, 123-140.

ANDERSON,J.R. Further arguments concerning representation for mental imagery: A response to Hayes-Roth and Pylyshyn. *Psychological Review*. 1979,86,395-406.

ATKINSON, R.C., HOLGREN,J.E. & JUOLA, J.F. Processing time as influenced by the number of elements in a visual display. *Perception and Psychophysics*. 1969,6,321-326.

BARLOW,R.E., BARTHOLOMEW,D.J. BREMNER,J.M. & BRUNK,H.D. *Statistical Inference under Order Restrictions: The Theory of Isotonic Regression*. New york: Wiley and Sons, 1972.

BERKSON,J. Application of the logistic function to bio-assay. *Journal of the American Statistical Association* 1944, 39,357-365.

BERKSON,J. A statistically precise and relatively simple method of estimating the quantel response, based on the logistic function. *Journal of the American Statistical Association*. 1953,48,565-599.

BLOCK & MARSCHAK, Random orderings and stochastic theories of responses. _in_: OLKIN, I. GLURYE, HOEFFDING, W. MADOW, W. & MANN, H. (Eds.) _Contributions to Probability and Statistics_. Standford University Press. 1960,97-132.

BOWER,G.H. Application of a model to paired-associate learning. _Psychometrika_, 1961,26,255-280.

BRADLEY, E. L. The equivalence of maximum likelihood and weighted least square estimates in the exponential family. _Journal of the American Statistical Association_. 1973,68,199-200.

BROADBENT,D.E. _Perception and Communication_. London: Pergamon Press, 1958.

BROADBENT and GREGORY. Stimulus set and response set: the alternation of attention. _Quaterly Journal of Experimental Psychology_, 16,1964,309-317

CHERNOFF, H. On the distribution of the likelihood ratio . _Annals of Mathematical Statistics_. 1954,25,573-578

CHERRY, C.E. Some experiments on the recognition of speech, with one and with two ears. _Journal of the Acoustical Society of America_. 25,1953,975-979.

COLLINS,A.M. and QUILLIAN,M.R. Retrieval from semantic memory. _Journal of Verbal Learning and Verbal Behavior_, 8,1969,240-247.

CRAIG,W. Replacement of auxiliary expressions. _Philos. Rev_. 1953,65,38-55.

DEUTSCH and DEUTSCH. Attention: Some theoretical considerations. Psychological Review. 70,1963,80-90.

DOUGLAS, G.A. Goodness of fit tests for paired comparison models. Psychometrika. 1978,43,129-130.

DROSLER J. Extending the temporal range of psychometric prediction by optimal linear filtering of mental test scores. Psychometrika. 1978, 43,533-549.

ESTES,W.K. Toward a statistical theory of learning. Psychological Review, 1950,57,94-107

ESTES,W.K. Statistical theory and the new "mental chemistry". Psychological Review. 1960,67,207-223.

FALMAGNE, J. C. Probabilistic choice behavior theory: Axioms as constraints in optimization. In N. J. Castellan & F. Restle (Eds.), Cognitive theory. New York: Lawrence Erlbaum Associates, 1978. Pp. 93-116.

FALMAGNE,J.C.,IVERSON G. and MARCOVICI S. Binaural "Loudness" Summation: Probabilistic Theory and Data. Psychological Review 86,1979,1,25-43.

GREENO J.G. & STEINER T.E. Markovian processes with identifiable states: General considerations and application to all-or-none learning, Psychometrika. 1964,29,309-333.

HENNING,H. Die Untersuchung der Aufmerksamkeit. in: ABDERHALDEN,E.: Handbuch der biologischen Arbeitsmethoden. Abt.vi, Teil B, Berlin: Urban &

Schwarzenberg, 1925, 593-802.

HEMPEL C.G. Fundamentals of concept formation in empirical science. in International Encyclopaedia Unified Science. vol 2, no. 7. University of Chicago Press, 1952.

HOLLAN, J.D. Features and semantic memory: Set-theoretic or network model? Psychological Review. 1975, 82, 2, 154-155.

JAMES, W. The principles of psychology. vol 1, 1890, Dover Publications, N.Y., 1950.

KAHNEMAN D. Attention and performance. 1973, Englewood Cliffs, New Jersey.

KEELE S. Compatability and timesharing in serial reation time. Journal of Experimental Psychology. 1967, 75, 529-539.

KOSSLYN, S. M. & POMERANTZ, J. R. Imagery, propositions, and the form of internal representations. Cognitive Psychology, 1977, 9, 52-76.

LAND, A.H. and POWELL, S. FORTRAN Codes for mathematical programming: Linear, Quadratic and Discrete. London: Wiley and Sons, 1973

LANSMAN, M. An attentional approach to individual differences in immediate memory. Technical report. Office of Naval Research, June 1978.

LEWIS J.E. Semantic processing of unattended messages using dichotic listening. Journal of Experimental Psychology.

1970, 85, 225-228.

LUCE D. and SUPPES,P. Preference, utility and subjective probability. in: LUCE,D.,BUSH,R. & GALANTER,E. (EDS) _handbook of mathematical psychology_, _vol III_ N.Y.Wley & sons,1965,249-410.

MANDLER,G. Consciousness: Respectable, useful, and probably necessary. _in_: SOLSO,R.L.(ED) _Information processing and cognition_: _The Loyola symposium_. Hillsdale ,New Jersey: Lawrence Erlbaum Associates,1975,229-254.

MEYER,D.E. On the representation and retrieval of stored semantic information. _Cognitive Psychology_. 1970,1,242-300.

MILLER G.A. The magical number seven, plus or minus two: Some limits on the capacity for processing inpormation. _Psychological Review_. 63,1956,81-97.

NAVON, D. & Gopher D. On the economy of the human-processing system. _Psychological Review_. 1979, 86,214-255.

NEYMAN, J. Contribution to the theory of the test. _In_: Neyman, J. (Ed) _Proceedings of the Berkeley Symposium on mathematical statistics and probability_. Berkeley: University of California Press, 1949

NORMAN D.D. and BOBROW,D.G. On data-limited and resource-limited processes. _Cognitive Psychology_. 1975,7, 44-64.

POSNER,M.I. and BOIES,S.J. Components of attention.

Psychological Review, 78 1971, 391-408.

PROCTOR, R.W. Attention and modality-specific interference in visual short-term memory. Journal of Experimental Psychology, Human Learning and Memory. 1978, 4, 3, 239-245.

PROCTOR, R.W. & FAGNANI, C. Effects of distractor-stimulus modality in the Brown-Peterson distractor task. Journal of Experimental Psychology. Human Learning and Memory. 1978, 4, 6, 676-684.

PYLYSHN, Z.W. What the mind's eye tells the mind's brain: A critique of mental imagery. Psychological Bulletin. 1973, 80, 1-24.

RAMSEY, F.P. The foundation of mathematics. London: Routledge & Kegan, 1931, 1965:4

RESTLE, F. The selection of stategies in cue learning Psychological Review 1962, 69, 329-343.

RIPS, L.J., Shoben, E.J. & Smith, Semantic distance and the verification of semantic relations. Journal of Verbal Learning and Verbal Behavior. 1973, 12, 1-20.

SCHEIBLECHNER, H. Specifically objective stochastic latency mechanisms. Journal of Mathematical Psychology. 1979, 19, 18-38.

SCHMALHOFER, F.J. Experimente zur gleichzeitigen Messung von sensorischer Beanspruchung und Aufmerksamkeit. Diplomarbeit, Regensburg University, 1978

SCHNEIDER, W. & SHIFFRIN, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review. 1977, 84,1-66.

SHIFFRIN,R.M. Capacity limitations in information processing, Attention and memory. in: ESTES Handbook of learning and cognitive processes. vol 4 Attention and Memory. LEA,1976

SHIFFRIN,R.M. & SCHNEIDER,W. Controlled and automatic human information processing: II Perceptual learning, automatic attending and a general theory. Psychological Review, 1977,84,127-190.

SCHULMAN,H. & GREENBERG,S. Perceptual deficit due to division of attention between memory and perception. Journal of Experimental Psychology 88,1971,171-176.

STERNBERG,S. High speed scanning in human memory. Science. 1966,153,652-654.

STERNBERG,S. & KNOLL,R. The perception of temporal order: Fundamental issues and a general model. in: KORNBLUM,S.(ed) Attention and performance iv N.Y. Academic Press,1973,629-685.

SUPPES,P. and ZINNES,J. Basic measurement theory. in: LUCE,D., BUSH & GALANTER (eds) Handbook of mathematical psychology. vol 1, N.Y. Wiley,1963.

TAYLOR, W.F. Distance functions and regular best asymptotical normal estimates. Annals of Mathematical

Statistics. 1953, 24,85-92.

TITCHENER,E. Lectures on the elementary psycholy of feeeling and attention N.Y. Macmillan,1908.

TOWNSEND, J.T. Issues and models concerning the processing of a finite number of inputs. in: KANTOWITZ,B.H. Human information processing: Tutorials in performance and cognition. Laurence Erlbaum Associates, 1974

TREISMAN,A. Contextual cues in selective listening. Quarterly Journal of Experimental Psychology 12,1960,242-248.

WILKS, S.S. The large sample distribution of the likelihood ratio for testing composite hypotheses. Annals of Mathematical Statistics. 1938, 9, 60-62.

WILKS, S.S. Mathematical Statistics. John Wiley & sons 1962

YAGLOM, A.M. An introduction to the theory of stationary random functions. Englewood Cliffs, Prentice Hall, 1962

YELLOT, J.I. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. Journal of Mathematical Psychology. 1977, 15, 109-144.

ZERMELO Die Berechnung der Tunierergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. Mathematische Zeitschrift. 1929, 436

FOOTNOTES

1) The author was supported by NSF grant BNS 77-06799 while writing this paper. The author is grateful to Walter Kintsch, Carol Nickerson, Peter Polson, and Gary McClelland for their comments on this paper.

2) The reported experiments are part of the author's thesis in partial fulfillment of the requirements of the Dipl. Psych. degree at Regensburg University. The author would like to thank the director of his thesis, Prof. Jan Drosler.

FIGURE CAPTIONS

Figure 1: A sample trial of the duration discrimination task as a function in time. The numbers 1 through 6 refer to the different signals used in this experiment.

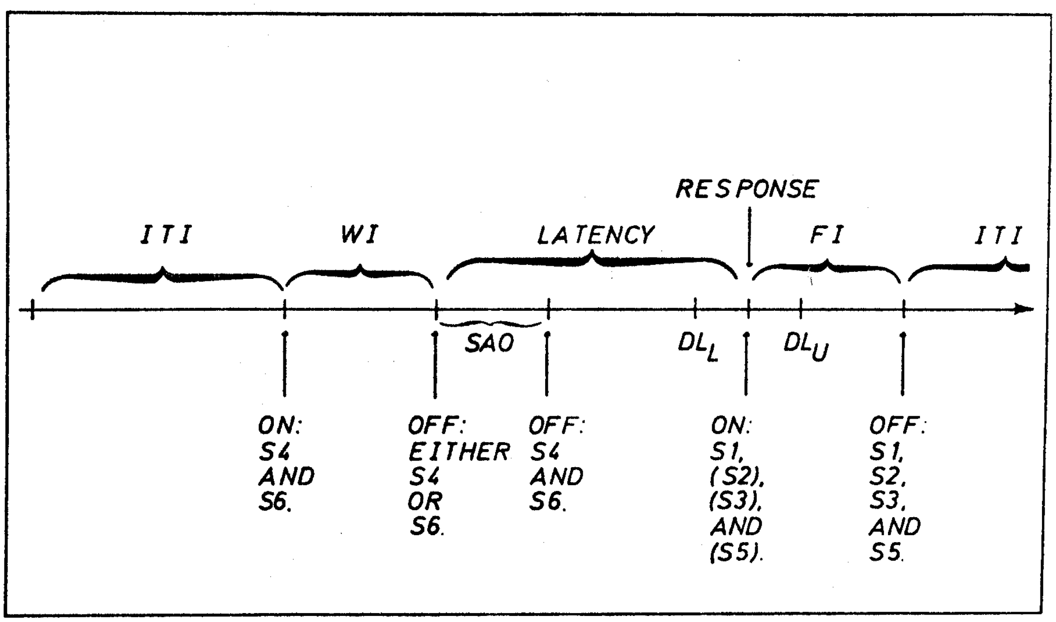

TABLE 1

Likelihood Ratio Tests for the Existence of a Single Solution Probability

	INDEPENDENT VARIABLES			TEST STATISTIC	
subject	pretraining and blocksize [# of trials]	SAO [msec]	DL_L-DL_U [msec]	-2 log λ	DF
1[a]	5500 150	60	200-400	12.32	9
2	5000 150	45	200-400	7.93	9
3	300 50	20 40 60 80 100	100-300 200-400 300-500 400-600	126.87*	19
4	3000 100	30 45 60 75	150-350 200-400 250-450 300-500	99.12*	15
5	2800 50	15 30 45 60 75 90	100-300 150-350 200-400 250-450 300-500	44.74	29
1[b]	3000 100	30 45 60 75	200-400 300-500 400-600	279.50*	11
7	3800 100	10 20 30	100-300 150-350 200-400 250-450 300-500	120.09*	14

For all subjects: ITI=1000 msec, WI=500 msec. Except for subject 3 (FI=800 msec), FI=500 msec.

a) b) Subject 1 participated in the experiment as a control and as an experimental subject.

*) The significance level is 0.10.

TABLE 2

Likelihood Ratio Statistics

of the Five Experimental Subjects

for the Three Theories T1, T2, and T3

	T1		T2		T3	
subject	$-2\cdot\log\lambda$	DF	$-2\cdot\log\lambda$	DF	$-2\cdot\log\lambda$	DF
3	1.26	1	23.33*	12	25.10*	12
4	15.78*	1	30.70*	9	28.67*	9
5	5.62*	1	22.69	20	71.56*	20
1b	0.75	1	17.86*	6	57.94*	6
7	2.36	1	21.50*	8	21.64*	8

*) Significant results are indicated by an asterisk.
The significance level was chosen to be 0.10.

SYMBOL IDENTIFICATION

1	the number one
l	the lower case letter "l"
α	the lower case Greek letter "alpha"
σ	the lower case Greek letter "sigma"
$\sum$	summation sign - the upper case Greek letter "sigma"
$\int$	the integration symbol
$\mathcal{A}$	the cursive upper case letter "A"
β	the lower case Greek letter "beta"
γ	the lower case Greek letter "gamma"
λ	the lower case Greek letter "lamda"
μ	the lower case Greek letter "mu"
θ	the lower case Greek letter "theta"
Δ	the lower case Greek letter "delta"
$\forall$	the "all" operator
$\exists$	the "existence" operator
$\mathcal{U}$	the cursive upper case letter "U"
$\mathcal{B}$	the cursive upper case letter "B"
$\mathcal{L}$	the cursive upper case letter "L"
$\in$	the lower case Greek letter "epsylon" meaning "element of"
φ	the lower case Greek letter "phi"
Φ	the upper case Greek letter "Phi"

ω	the lower case Greek letter "omega"
Ω	the upper case Greek letter "Omega"
χ	the upper case Greek letter "Chi"