Sperling, G., and Dosher, B. A. (1986). Strategy and optimization in human information processing. In K. Boff, L. Kaufman, and J. Thomas (Eds), *Handbook of Perception and Performance. Vol. 1.* New York: Wiley. Chapter 2, Pp. 1-65.

# HANDBOOK OF PERCEPTION AND HUMAN PERFORMANCE

## VOLUME I
## Sensory Processes and Perception

*Editors:*

**KENNETH R. BOFF** *Armstrong Aerospace Medical Research Laboratory*

**LLOYD KAUFMAN** *New York University*

**JAMES P. THOMAS** *University of California at Los Angeles*

**A Wiley-Interscience Publication**

**JOHN WILEY AND SONS**

**New York • Chichester • Brisbane • Toronto • Singapore**

Sperling, G. and Dosher B. A. (1986). Strategy and optimization in human information processing. *In* K. Boff, L. Kaufman, and J. Thomas (Eds.), *Handbook of Perception and Performance. Vol. 1.* NY: Wiley. Chapter 2, Pp. 1-65.

# CHAPTER 2

# STRATEGY AND OPTIMIZATION IN HUMAN INFORMATION PROCESSING

**GEORGE SPERLING**
Department of Psychology, New York University, New York, New York

**BARBARA ANNE DOSHER**
Department of Psychology, Columbia University, New York, New York

CONTENTS



The authors wish to dedicate this chapter to the memory of Marilyn Shaw (1946–1983).

This chapter deals with the recent methods and theories developed to study subjects' strategies in the performance of perceptual and cognitive tasks. *Strategy* refers to the selection and sequencing (by the subject) of mental operations in the performance of a task. In some cases, important components of the strategy are externally observable, as in the sequence of eye movements that a subject makes in searching a display for a target signal. In the cases considered here, the strategy is not directly observable but must be inferred by means of elaborate theories from repeated, indirect observations. A guiding principle in these theories is that the subject chooses a strategy with the intent of optimizing behavior with respect to the external situation, given his or her internal limitations. For example, the sequence of eye movements in search cannot be understood unless we know critical situation-determined facts (such as where the subject expects to find targets and what the payoffs for various outcomes of the search task are) and critical internal constraints (such as the distribution of acuity around the point of fixation and the maximum rate of eye movement).

It is much more difficult to establish the existence and to provide a quantitative description of an unobservable strategy than of a directly observable one. Consequently, this chapter is restricted to relatively simple tasks performed in very brief time periods. Nevertheless, the same general principles of optimization that apply to decision making and resource allocation on a large scale apply to "micro" strategies that govern decision making and mental resource allocation on a millisecond time scale. In psychology, signal detection theory (SDT) introduced the concepts of optimization to the most elementary perceptual tasks, simple detection of threshold signals. In SDT, strategy is restricted to variation of a decision criterion (the bias) to maximize the subject's expected utility. In earlier detection theories, the subject's strategy consisted of selecting a guessing algorithm applied on some fraction of trials, usually those on which the subject had acquired insufficient information for a stimulus-controlled response. In this chapter, the scope of all these theories is expanded. It is assumed that the goal of the subject is to maximize rewards (utility). In SDT, this optimization principle is explicit. Nevertheless, optimization has been obscured by particularizing the optimization principles inherent in SDT to signal detection so that some general properties and problems of optimization, such as iso-utility contours, complex payoff rules, multidimensional strategies, and changeover costs, have been overlooked or neglected.

Subjects' strategies in SDT and, more generally, in compound tasks are *decision* strategies, governed by the same general principles that govern decision making under uncertainty, that is, decision making with partial or inaccurate information. On the other hand, in other multitask, divided-attention situations, such as driving a car and listening to the radio, subjects' strategies deal with the allocation of mental processing resources. Remarkably, the optimization theory underlying resource allocation is largely isomorphic to the theory underlying decision making in detection! This is a powerful argument for considering these two areas of study together.

Reaction-time experiments are another domain in which subjects' performance can only be understood in terms of their unobservable strategy. Strategy manifests itself, for example, in the choice of a particular operating point on a speed–accuracy trade-off; the theory of how this choice is made is an extension of the optimization theory governing detection and resource allocation.

The methods for measuring observable motor responses have been developed to a high degree of precision. New experimental methods allow equally precise measurement of the reaction time of an observer's attention shift. This is the precise characterization of the dynamics of the mental resource allocation involved in strategies. Finally, some classical psychophysical procedures, and some new ones, are analyzed in terms of how they control, or fail to control, subjects' strategies and what this means in terms of choosing experimental procedures appropriate for the subsequent uses of the data.

# 1. CONCURRENT VERSUS COMPOUND: A TASK TAXONOMY

In the study of signal detection, attention, and performance, one is often interested in the subject's ability to perform several tasks (or to respond to several classes of stimuli) in the same situation. While there are many ways to combine subtasks into a combined task, two kinds of combinations are by far the most common: *concurrent* combinations and *compound* combinations (Sperling, 1984). The exact manner in which the component tasks are combined is critical for interpreting the results and for understanding the subject's strategy. This section contains the formal definitions of a task and of compound and concurrent combinations, with some representative examples. The interpretation of results is considered in Section 4, after the issues of optimization have been considered.

## 1.1. Task Definition

The discussion here is restricted to *discrete* tasks, tasks which consist of discrete trials, as contrasted to *continuous* tasks. One trial of a discrete task consists of the presentation of a stimulus $s$ and the observation of a response $r$. For example, a visual search task might present a $5 \times 5$ array with the digit *1* in one of the 25 locations. The task is to report location. One characterization of this is a 25-element stimulus set and a 25-element response set. The *outcome* of a trial is the $s,r$ pair resulting from that trial. For every outcome, there is assumed to be a utility, which is a real number representing the value of that outcome to the subject.

The description of a discrete task can be made rigorous (Sperling, 1983). The task is a triple $(S, R, U)$, consisting of two sets (*stimuli, responses*) and a function, the utility function. Let $S$ denote the set of alternative stimuli in a task. A particular stimulus is represented as $s$ ($s \in S$, that is, $s$ is an element of $S$). Let $R$ denote the set of alternative responses, and let $r$ ($r \in R$) denote a particular response. Let the symbol $X$ denote the Cartesian product. Then the utility function $U$ is a mapping of $S X R$ into the real numbers Re, that is, $U: S X R \rightarrow \text{Re}$. Ideally, in carefully controlled procedures, utility is defined explicitly by the experimenter. Commonly, utility is not fully defined by the experimenter, and the subject is assumed to define a utility function implicitly. The utility function may be a simple two-valued function (e.g., 0,1 to represent "wrong," "right"), or it may be a complicated real-valued function that involves, for example, the speed and accuracy of choice reaction times.

## 1.2. Compound Tasks

A compound task combines two or more component tasks in such a way that each trial consists of a single stimulus drawn randomly from one of the component tasks and a response, also from one of the component tasks. A compound task $(S,R,U)$ is a combination of component tasks $(S_i,R_i,U_i)$ that satisfies certain conditions. Consider a particular component task $i$. Let $S_i$ represent the set of alternative stimuli in task $i$; let $R_i$ represent the set of alternative responses; and let the real-valued function $U_i(s_i,r_i)$ represent the utility of a stimulus–response pair $(s_i,r_i)$. (Note that subscripts are used to denote a task; lowercase letters are used to indicate a stimulus within a task; and $s_i$ refers to *any* element of $S_i$, the stimulus set of task $i$; i.e., the subscript refers to the task, not to a particular stimulus in the task.)

**1.2.1. Definition.** The task $(S,R,U)$ is a compound task composed of component tasks $(S_i,R_i,U_i)$, $1 \leq i \leq n$, $n \geq 2$, if and only if the following three conditions hold:

*Condition (CP-1)*

$$S = \bigcup_{i=1}^{n} S_i ,$$

The notation $\cup_{i=1}^{n} S_i$ means union of $S_i$; that is, every stimulus that can occur in any task $i$ occurs in the union. The stimulus presented on each trial of the compound task is a selection of one stimulus from any one of the component tasks.

*Condition (CP-2)*

$$R = \bigcup_{i=1}^{n} R_i ,$$

Any response that can occur in a component task can also occur in the compound task. Note that it is possible for new stimulus–response pairs to occur in the component task, since a stimulus from component task $i$ might elicit a response from component task $j$. In defining utility $U(s_i,r_j)$, there are two cases. In the first case, the response to a stimulus from task $i$ is a member of $R_i$. In this case, utility is simply proportional to the utility of that stimulus–response pair within task $i$ alone. The relative importance of task $i$ in the task ensemble is expressed by the positive constant $\alpha_i$ in Condition (CP-3a).

*Condition (CP-3a)*

$$U(s_i,r_i) = \alpha_i U_i(s_i,r_i), \qquad \alpha_i > 0, \qquad 1 \leq i \leq n .$$

The second case deals with utility when subjects respond to a stimulus $s_i$ from one of the component tasks with a response $r_j$ from another.

*Condition (CP-3b).* The utility of response $r_j$ to stimulus $s_i$, $U(s_i,r_j)$, $i \neq j$, is inversely related to the *distance* of response $r_j$ from an optimal response $r_i^*$ to the particular $s_i$; this utility is expressed as a utility function, $V(r_j,r_i^*)$, defined on two responses. The utility of a wrong response cannot exceed the utility of the optimal response, for example, $U(s_i,r_i^*) \geq V(r_j,r_i^*)$. Alternatively, utility might have been defined by the distance between stimuli.

**1.2.2. Examples.**

***1.2.2.1. Visual Search.*** Consider the visual search example described in the previous section in which the task consisted of locating the digit *1* in a 5 × 5 array of characters. This task is combined with that of locating a *2* in a 5 × 5 array. Either a *1* or a *2* (but not both) appears unpredictably on each trial. The way to conceptualize the compounding operation is to imagine two urns: Urn 1 contains all possible stimuli consisting of the target "1" and 24 nontarget letters; Urn 2 contains all possible stimuli with target "2" and nontarget letters. A compound task trial consists of selecting a stimulus from either Urn 1 (task 1) or Urn 2 (task 2).

Searching for the location of *1* or *2* is a *compound* task that might be used to study the question of attentional limitations of searching for two characters at the same time. This is a special case of a compound task where the response sets (25 locations) of the two component tasks are identical. In this case, many of the complexities of defining utility are avoided, and Conditions (CP-3a) and (CP-3b) reduce simply to

*Condition (CP-4)*

$$\text{If } R = R_i = R_j \quad \text{for } 1 \leq i,j \leq n \ ,$$

$$U(s_i, r) = \alpha_i U_i(s_i, r) \qquad \alpha_i > 0 \ .$$

***1.2.2.2. Signal Detection.*** Another example of a compound task is a signal detection task in which one of a number of stimuli (i.e., one of four tones of different frequencies) is presented unpredictably on each trial, and the response is saying either "signal" or "noise." The discrimination from noise of each tone separately can be considered to be a component task, so $S_i = (t_i,n)$, $1 \leq i \leq 4$; $R_i = R =$ ("signal," "noise") for all $i$, and this example is another of the special cases covered by Condition (CP-4).

Signal detection illustrates a potentially confusing aspect of compound tasks. The selection of the stimulus value noise (also called *no signal, null*) from Task 1 produces precisely the same physical stimulus event as the selection of noise from Task 2. Two conceptually different events (presentation of stimuli from Task 1 or Task 2) have the same physical instantiation. This is not a problem unless the subject is required to discriminate them, which obviously was not required in this example. The discrimination problem does manifest itself when the signals in the subtasks (e.g., Tone 1, Tone 2) are not completely discriminable, and we have to explicitly consider separately the tasks of *detection* and of *discrimination at threshold*, a subject about which much has been written.

***1.2.2.3. Choice Reaction Time.*** Finally, consider a choice reaction-time experiment where the subject is presented with one of five lights and must press one of five corresponding keys. This choice task may be thought of as a compound combination of five simple reaction-time tasks.

A simple reaction task consists of a warning tone (which can be regarded as part of the experimental situation, much like the chair the subject sits on) followed, after a variable foreperiod, by a reaction signal. A stimulus, then, consists of a foreperiod and a signal. The response is a reaction *time*. Each component task has associated with it a set of stimuli; the compound task consists of a selection of one stimulus from one of the component tasks, followed by a response. The response to a stimulus from set *i* may be with the finger that is appropriate to component task *j*, the situation to which Condition (CP-3b) applies. The very large observed increase in reaction time with an increase in the number of alternatives points out the importance of uncertainty in the analysis of compound reaction-time tasks. This issue is considered in Section 5.

## 1.3. Concurrent Tasks

A concurrent task is one that combines two or more tasks in such a way that each component task must be performed on each trial. This is generally exemplified by situations where each component task is performed independently, like driving an automobile while listening to the news on the radio. A concurrent task $(S,R,U)$ is a combination of component tasks $(S_i, R_i,U_i)$ that satisfies the following conditions.

Let $S_i$ represent the set of alternative stimuli in task $i$, $s_i \in S_i$. Let $R_i$ represent the set of alternative responses, $r_i \in R_i$. Let the utility $U_i$ of a particular stimulus–response pair be a real number defined for every pair $(s_i,r_i)$.

**1.3.1. Definition.** The task $(S,R,U)$ is a concurrent combination of the $n$ component tasks $(S_i,R_i,U_i)$, $1 \leq i \leq n$, $n \geq 2$, if and only if the following four conditions hold:

*Condition (CC-1)*

$$S = S_1 X\, S_2 X \ldots X\, S_n$$

*Condition (CC-2)*

$$R = R_1 X\, R_2 X \ldots X\, R_n$$

*Condition (CC-3)*

$$U(s,r) = U[(s_1,s_2, \ldots, s_n),(r_1,r_2, \ldots, r_n)]$$

$$= H[U_1(s_1,r_1), U_2(s_2,r_2), \ldots, U_n(s_n,r_n)]$$

where $H$ is strictly increasing in each variable.

*Condition (CC-4).* The stimulus components $s_i$ from each component task of the concurrent task are chosen independently of each other.

**1.3.2. Explanation.** We consider first Condition (CC-1). The concurrent stimulus $S$ is regarded as an $n$-dimensional vector whose components $S_i$ are the stimuli of the component tasks. Suppose *i* and *j* are two component tasks of the concurrent combination. Condition (CC-1) means that any stimulus that can occur in task *i* can occur in combination with any stimulus that can occur in task *j*. Condition (CC-4) asserts that this co-occurrence is independent.

Condition (CC-2), the possible co-occurrence of any response from task *i* with any other response from task *j*, parallels exactly Condition (CC-1) with stimuli.

Condition (CC-3) asserts (1) that the utility mapping for a component task does not change when that task occurs concurrently with other tasks and (2) that the utility of the concurrent combination of tasks is an increasing function of the utility of each component task. The function $H$ is stated in very general form. One commonly used function is a weighted linear combination of component utilities:

*Condition (CC-5)*

$$H(U_1, U_2, ..., U_n) = \sum_{i=1}^{n} \alpha_i U_i, \quad \alpha_i > 0 \ .$$

Given Condition (CC-5), an equal attention combination is one where $\alpha_i = \alpha_j$ for all $i, j$.

Another common, nonlinear, utility function is the logical product (or intersection) exemplified by a high school curriculum. In this example, a component task is an individual high school course. The examination questions are the stimulus; the student's answers are the response. The utility is 1 if the course is passed, 0 if failed. The concurrent task consists of taking all the courses necessary to graduate. Graduating is possible only if all the component courses are passed. Thus:

*Condition (CC-6)*

$$H(U_1, U_2, ..., U_n) = 1 \quad \text{iff} \quad \bigcap_{i=1}^{n} U_i = 1$$

$$= 0, \ \textit{otherwise} \ .$$

Of course, this particular example is unrealistically simple. However, it illustrates that there is nothing in the definition of a concurrent task that requires the component stimuli to be presented at precisely the same time (though this is the usual case), merely that one stimulus from each task be presented on each trial, even when the trial lasts for years.

**1.3.3. Examples.** In a numeral detection task, $S_1$ represents the presence of numeral 1 in some location $l_1$, $l_1 \in L_1$, and $S_2$ represents numeral 2 in a location $l_2$, $l_2 \in L_2$. Concurrency requires that, in the concurrent task, every combination $L_1 X$ $L_2$ of $l_1$ and $l_2$ can occur. In particular, consider the case where $L_1 = L_2$, and $L_1$ represents the locations in a $3 \times 3$ array. In the concurrent task, not only must targets 1 and 2 both occur, but by Condition (CC-4) they must also occasionally occur in the same location. This makes psychological sense if the two targets occur at different times (e.g., in successive arrays) but not when they occur in the same location at the same time. However, concurrent search is possible within an array. For example, Sperling and Melchner (1976b, 1978a) described a visual search task in which subjects concurrently detected the location of digit targets in both an outer and an inner array. Since the location sets for these two tasks did not overlap, these two detection tasks occurred concurrently on the same stimulus frame.

Some examples of concurrent tasks that have been studied experimentally are: shadowing one auditory message (repeating it with as little delay as possible) while attempting to listen to and recall another (Glucksberg & Cowen, 1970; Treisman, 1964), shadowing one message while sight-reading and playing a piano score (Allport, Antonis, & Reynolds, 1972), recalling digits heard in the left ear concurrently with digits heard in the right ear (Broadbent, 1954), and reporting concurrently on the presence or absence of three independent near-threshold tones—500, 810, and 1320 Hz (Sorkin, Pohlman, & Woods, 1976) or two independent visual spatial frequencies (Hirsch, Hylton, & Graham, 1982).

**1.3.4. Overview.** From the formal description of concurrent and compound tasks, it should be clear that both types of task have been used widely in the investigation of signal detection, attention, decision processes, and other aspects of performance, although compound tasks are perhaps the more common. Difficulty in interpreting compound tasks arises because of an inevitable *signal uncertainty* component. An ideal detector would show a loss in a compound task, and a quantitative model of the uncertainty decrement is necessary to determine whether there is an attentional loss (due to an insufficiency of processing resources) in addition to the uncertainty loss. Results of concurrent tasks are interpretable directly as determined by human limitations, but difficulties may occasionally occur in the interpretation of loss in performance of concurrent tasks as due to attentional limitations, reporting bottlenecks, or other limitations. Both kinds of tasks may be required to answer a particular question. The advantages and disadvantages of these paradigms are discussed in subsequent sections.

# 2. MAXIMIZING UTILITY: AN ATTENDANCE EXAMPLE

This section introduces the classroom attendance example (Sperling, 1984; Sperling & Melchner, 1978a, 1978b), which is a model for resource theories of attention. The concept of an *operating characteristic* is introduced and related to the notions of *utility* and *utility maximization* as a determinate of the subjects' *strategy*.

## 2.1. Information Densities and the Performance Operating Characteristic

A student wishes to attend two classes, A and B. The classes are offered in adjacent classrooms, so that going from one class to the other requires a negligible amount of time. Once the student leaves a classroom, return to it is not permitted. At the end of the term, the student takes an examination in each class. On the examinations, each instructor asks one question from each lecture. It is assumed that the tested information is distributed uniformly over the lecture period.

The left panels of Figure 2.1 show three examples that differ in the overlap of times at which the classes are offered. Cases (a), (b), and (c) differ in the degree of competition for attendance. The student's only strategic option is the criterion time c at which he moves from class A to class B. Depending on the student's choice of c, he or she can control the performance level in the class A and class B examinations. The joint performance in the two classes (probability correct on examination questions) is plotted on the right panels of Figure 2.1. The joint performance at various switching times c defines the *attendance operating characteristic*. As the class times overlap less and less, it is possible to choose c to result in better and better performance, so that in the case in Figure 2.1(c) the student can perform perfectly in both classes by choosing $c = 3$ o'clock as the switching time. This level of joint performance represents the independence point in Figure 2.1(c); the student performs as well in each of the concurrent tasks as in the isolated component tasks. The independence point is achievable only in the nonoverlapping case, where there is no competition for the student's attending resources.

When class times overlap, it is not possible to achieve a score of 100% in both classes. In case (b), the student can perform

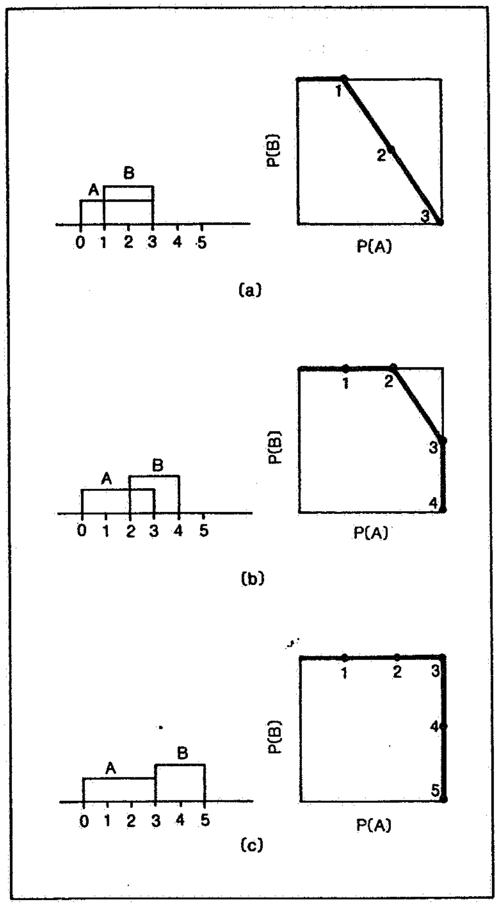



**Figure 2.1.** Classroom information densities and the corresponding attendance operating characteristics (AOCs). The left panels show hypothetical information densities as a function of time in hours for two classes, *A* and *B*, with different, uniform density functions. The right panels show the resulting AOCs with points indicating several classroom switching times. The axes of the AOCs represent performance (the probability of correct responses on examinations) in the two classes. (a) A case where class *B* overlaps completely with *A*, and the information density per unit time is higher for *B*. (b) An intermediate degree of overlap. The resulting AOC is closer to the upper right corner (the *independence point*) indicating less processing conflict. (c) A case of no overlap and the resulting AOC.

perfectly in class A and 50% in class B, perform perfectly in class B and 66.7% in class A, or do something in between. Or the student could do something perverse, such as sit in classroom B until 2:00 P.M. and then switch to A or not attend either class. Perverse and obviously inferior strategies are not considered in this chapter. The next section examines how the student *should* choose among the reasonable strategies.

## 2.2. Utility and Strategy Selection

For the student to choose a particular strategy from the set of reasonable strategies, the utility of all the candidate strategies must be determined. The method is illustrated for the situation in Figure 2.1(b).

Suppose that classes A and B contribute equally to the student's overall grade point average (the higher the average, the greater the utility), and no other considerations enter into this computation. The corresponding utility function is

$$u(x_A, x_B) = 50(x_A + x_B) \tag{1}$$

where $x_i$ is the probability of a correct response in class $i$ and $u(x_A, x_B)$ is utility. (Because utility is known only up to an arbitrary, strictly increasing monotonic transformation, scale factors are introduced only to clarify the example.)

The optimal strategy is to attend all of class B and as much of class A as possible, thereby achieving an average of 83.3%. This and other implications of the particular utility function are made intuitively obvious by plotting iso-utility contours together with the operating characteristic, as in Figure 2.2. This iso-utility, or indifference, curve approach, according to Due (1951), was first suggested by Pareto (1909) and later developed by Hicks and Allen (1934).

Given an explicit utility function, the utility of each strategy can be computed directly. That is, utility can be written as a function of the class-switching time $c$ by writing the examination scores as a function of $c$:

$$u(c) = 50\left[\frac{\min(c,\ 3)}{3} + \frac{\min(4-c,\ 2)}{2}\right], \qquad 0 \leqslant c \leqslant 4 \tag{2}$$

where $\min(x,y)$ is defined as the smaller (minimum) of $x$ and $y$, and $c$ is measured in hours (with noon taken as 0).

The parallel, diagonal lines in Figure 2.2(b) represent *iso-utility contours*. Utility can be computed for every point in the joint performance space $(x_A, x_B)$ whether or not that point is achievable. The parameters used to label iso-utility contours in Figure 2.2(b) indicate their utility. The *attending operating characteristic* crosses iso-utility contours until it touches (is tangent to) the maximum utility contour it can reach. The highest contour reached is 83.3%; this occurs with a class-switching time of 2:00 P.M.; it results from a perfect score in class B and 66.7% in class A. The reason for the relative neglect of class A is that useful information has higher density per unit time in class B than in class A, and therefore the *marginal utility* of attending class B is greater than that of attending class A. The student should exchange time in class A for time in class B whenever possible.

Suppose that the utility of success in class B were only 2/3 that of class A; this would happen if class A were weighted as three credit-hours and class B as two credit-hours. Then

$$u(x_A, x_B) = 60(x_A + 2/3x_B) \ . \tag{3}$$

These iso-utility contours are shown in Figure 2.2(c). In this case, the lower utility of class B exactly offsets its higher-information density, and a switching time anywhere between 2 and 3 will maximize utility as defined by Eq. (3).

In a third case, suppose that what matters is not grade point average but simply passing all the courses. The utility is 1.0 if all courses are passed, 0 otherwise. Figure 2.2(d) illustrates utility graphs for three minimum-required passing grades, 50%, 65%, and 80%. The curves in Figure 2.2(d) are not iso-utility contours as before, but divisions of the graph into two regions: (1) pass both courses and (2) fail one or both courses. For convenience, the three boundaries under consideration are represented on one graph. All the reasonable strategies suffice when the minimum passing grade is 50%; about half of the reasonable strategies are adequate with a minimum passing grade of 65%; only one strategy will achieve 80%, which is the highest grade simultaneously achievable in both courses. To achieve a grade of 80%, the student attends 80% of each class;

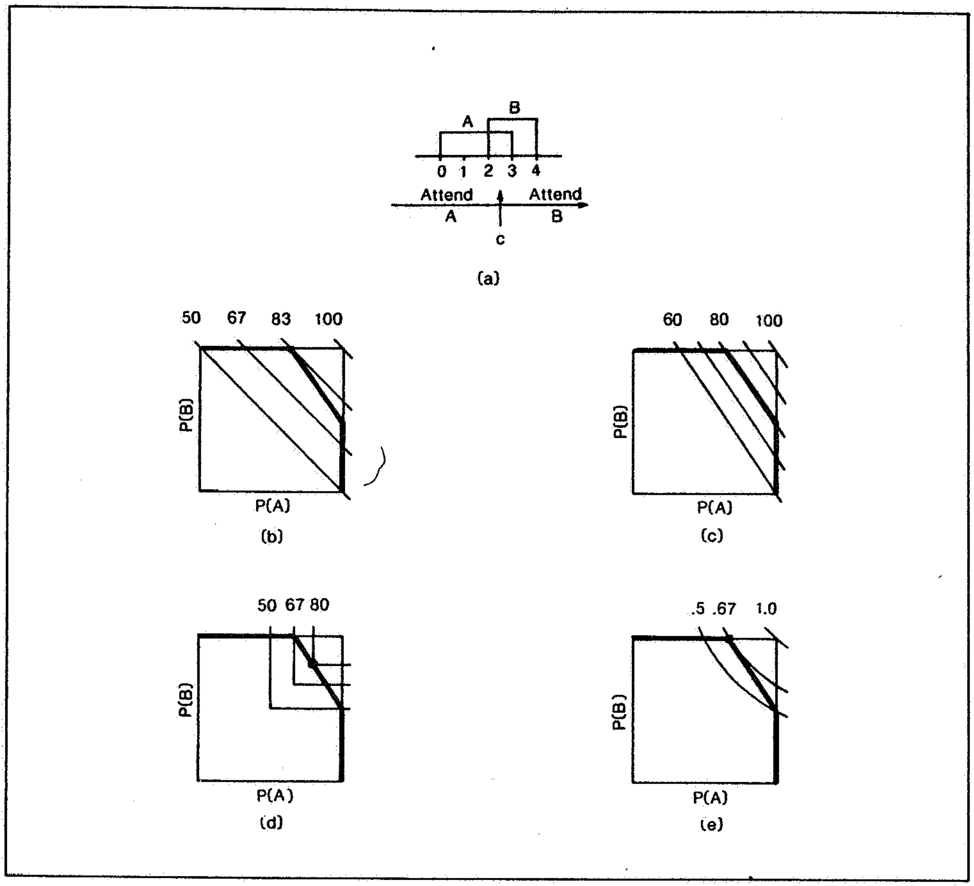



**Figure 2.2.** A classroom attendance example with four different utility functions. (a) Two partially overlapping class information density functions. (b) The resulting attendance operating characteristics (AOCs) with iso-utility contours for an equal weighting utility function, $u(x_A, x_B) = 50(x_A + x_B)$, where $x_i$ represents performance in class $i$ ($i = A, B$). Several iso-utility contours are shown; the parameter indicates their relative value. Optimal performance occurs where the AOC intersects or is tangent to the highest possible iso-utility contour. The intersection corresponds to a switching time of c = 2.0 in this example. (c) The same AOC with a utility function that weights performance in class A and B directly in accordance with the relative durations of the two classes—a 3:2 ratio: $u(x_A, x_B) = 20(3x_A + 2x_B)$. Because the iso-utility contour coincides with one branch of the AOC, there is an extended range of optimal switching times. Time in class A and class B is precisely equal in value. (d) A pass-fail utility function. The student requires a given score (e.g., 80%) to pass each class and needs to pass each class to graduate. This defines iso-utility *graduation regions* above and to the right of the lines shown, and *fail regions* elsewhere. Performance everywhere in a region is equivalent, since the pass-fail utility rule does not value increments (or decrements) in performance above (or below) the pass-fail criterion, nor does it discriminate between failing one or failing more courses. Criterion c in panel (a) represents the optimal switching time (c = 2.4). The iso-utility contours exhibit the property of diminishing returns to an extreme degree; attendance beyond that required for a passing grade produces no additional return whatsoever. (e) A case with statistical uncertainty in the probability $P(i)$ that a class $i$ will be passed, such that the probability of passing both classes is $P(A\&B) = P(A)P(B)$. Utility is assumed to be directly proportional to $P(A\&B)$. The utility function favors equidistribution of effort; however, in this example it cannot overcome the greater utility of time spent in class B. (d. is from G. Sperling, A unified theory of attention and signal detection, in R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

that is, he or she switches from class A to B after 2.4 hours in class A (at 2:24 P.M.). Ironically, this strategy, which is the only one that will enable the student to pass both courses when a passing grade of 80% is required, is the strategy that would cause the person to *fail* both courses when a passing grade of 80.1% is required.

The iso-utility contours in the pass-fail example reflect the important principle of *diminishing returns*. Attendance up to the point of achieving a passing grade is valuable; additional attendance offers no return whatever.

A different example than simple pass or fail is constructed by assuming that the probability $p(i)$ of passing a course $i$ is simply proportional to the cumulative attendance $x_i$ in the course. As in the previous examples, $p(i) = x_i/\tau_i$, where $\tau_i$ is the number of classroom hours. The event that determines utility is passing both courses. The probability, $p(A\&B)$, of passing both class A and B is $p(A\&B) = p(A)p(B)$. Iso-probability contours for $p(A\&B)$ (which also represent iso-utility contours) are shown in Figure 2.2(e).

If it is assumed that utility is directly proportional to $p(A\&B)$, then the iso-utility contours of Figure 2.2(e) do not exhibit diminishing returns. For a fixed value $x_B$ of attendance in class B, an additional minute ($\Delta x_A$) of attendance in class A produces precisely the same increment in $p(A\&B)$, independent of the value $x_A$ has already attained. However, this utility function does represent a high value for *equidistribution of effort.* It represents the case where it is advantageous for the student to be well rounded, that is, to have approximately equal scores (as high as possible) in all classes, or the case of the professor who is better off with a reasonable level of knowledge about the subject matter *and* a reasonable level of communicative ability than with a surplus of one and an insufficiency of the other. While this utility for equidistribution of effort would favor actual equal distribution if the marginal utility of time spent in each class were equal, it cannot in this example overcome the greater utility of time spent in class B, and the optimum is the same as in Figure 2.2(b).

## 2.3. Interpretation of the Classroom Example

The classroom example of Section 2.2 is a model for the case of attention to two *concurrent* tasks. The temporal overlap of the courses is analogous to an overlap in requirements for processing resources (Navon & Gopher, 1979). Processing resources are considered in detail in Sections 3.4 and 3.7. The subject's strategic choice in attentional tasks involves the allocation of mental resources (i.e., attention) between the two tasks.

To avoid confusion, it is very useful to use different words for real-world quantities and for the variables that represent them in a theory or model. "Attention" normally is used as a real-world term; "processing resources" is a mathematical construct in models proposed to account for attentional phenomena. Insofar as a particular model is in one-to-one correspondence with some set of attentional phenomena, there may be (but need not be) an observable or potentially observable quantity, such as "amount of attention," that corresponds to the amount of processing resources. The classroom model is a single-processor model, in which *time* is the resource divided between the processes. However, the classroom model generates performance operating characteristics that describe data perfectly well even when time is not the critical resource, for example, when memory capacity is critical. Thus the hours from noon to 5:00 P.M. could represent five memory slots that could be allocated to storing items either from list A or from list B according to the task demands and conditions. The essential aspect of the classroom model is the computational procedure and not the particularities of the classroom. In the resource analogy to attention, the performance operating characteristic is called the *attention operating characteristic* (Sperling & Melchner, 1978b; see also Sperling, 1984, p. 112). When utility functions are not defined explicitly, the subject performing concurrent tasks is assumed to choose a strategy according to subjective utilities. Because subjective utilities are difficult to discover and because they are likely to be labile, the experimenter should endeavor to make the utility function as explicit as possible.

# 3. SIGNAL DETECTION THEORY, ATTENTION, AND ECONOMICS

The formal analogy of the classroom attendance example to models of signal detection, attention, and economic production is developed here with special emphasis on the different interpretations of the basic components of the model in these various contexts. The operating characteristics of the classroom example are related to the receiver operating characteristics of signal detection theory, the production possibilities frontiers of economics, and the attention operating characteristics of human information processing. The approaches to optimization in these domains are outlined.

## 3.1. Attendance Example Generalized

In this section, the logic of the (classroom) attending operating characteristic is reviewed in a more formal way. During class period $i$ and at the moment in time $x$, let the rate at which an instructor is presenting information be given by $p_i(x)$. The information rate $p_i(x)$ is assumed to be 0 when class is not in session and to be nonnegative while class is in session. The total *amount of information* $E_i$ presented in Class $i$ [$E_i = \int_{start}^{finish} p_i(x)dx$, $i = 1,2$] is assumed to exist and be bounded. For the attending strategy in which a student attends class 1 from its start until time $c$ and then attends class 2 until its finish, the amount of information $E_i$ accumulated in each class is given by:

$$E_1 = \int_{-\infty}^{c} p_1(x)dx \quad \text{and} \quad E_2 = \int_{c}^{\infty} p_2(x)dx \,. \tag{4}$$

Information accumulates only from the starting time $x_0$ of the class. However, since $E_1(x)$ is zero for $x < x_0$, it is convenient to write the integral from $-\infty$ to $c$ rather than from $x_0$; similarly, for class 2. The performance operating characteristic is a graph of $E_2$ versus $E_1$ as $c$ varies.

It will be convenient in the following discussion to consider the special case of the classroom attendance example where the information densities are approximately normal (and the class periods are equivalent). Thus instructors take a while to warm up before they reach their maximum exposition rate, and having once reached this rate, they quickly begin to tire, and less information is presented late in the class period. Figure 2.3(a) illustrates the information rates estimated for two instructors, one in a nursing class given from noon until 1:40 P.M., the other in a Spanish class given from 12:20 to 2:00 P.M. Figure 2.3(b) illustrates the attendance operating characteristic for a student who attempts to take both classes. With these normal distribution assumptions, the sharp edges of the previous performance operating characteristics have been rounded to a smooth curve, but the logic remains the same.

## 3.2. Signal Detection Theory

For simplicity of exposition, consider now a case of signal detection theory (SDT) that is more particular than it needs to be, but which can readily be generalized. An experimenter presents two kinds of trials: those on which only noise $N$ is presented and those on which a signal plus noise $S + N$ is presented. (The notation $S + N$ and $S$ will be used interchangeably when the addition of noise is irrelevant or obvious from context.) The observer's task is to distinguish between these two kinds of trials, that is, to say "signal" or "yes" whenever believing $S$ was presented and to say "noise" or "no" otherwise. The possible outcome of a signal trial is either a correct detection (i.e., hit) or a failure to detect (i.e., miss); the possible outcome of a noise trial is either a correct rejection or a false detection (false alarm).

Signal Detection Theory proposes that any stimulus (either $S$ or $N$) is represented internally by a real number $x$ along an internal continuum. This real number is regarded as a random variable, with conditional distributions (given $N$) $p_N(x)$ and (given $S$) $p_S(x)$ usually assumed to be normally distributed.

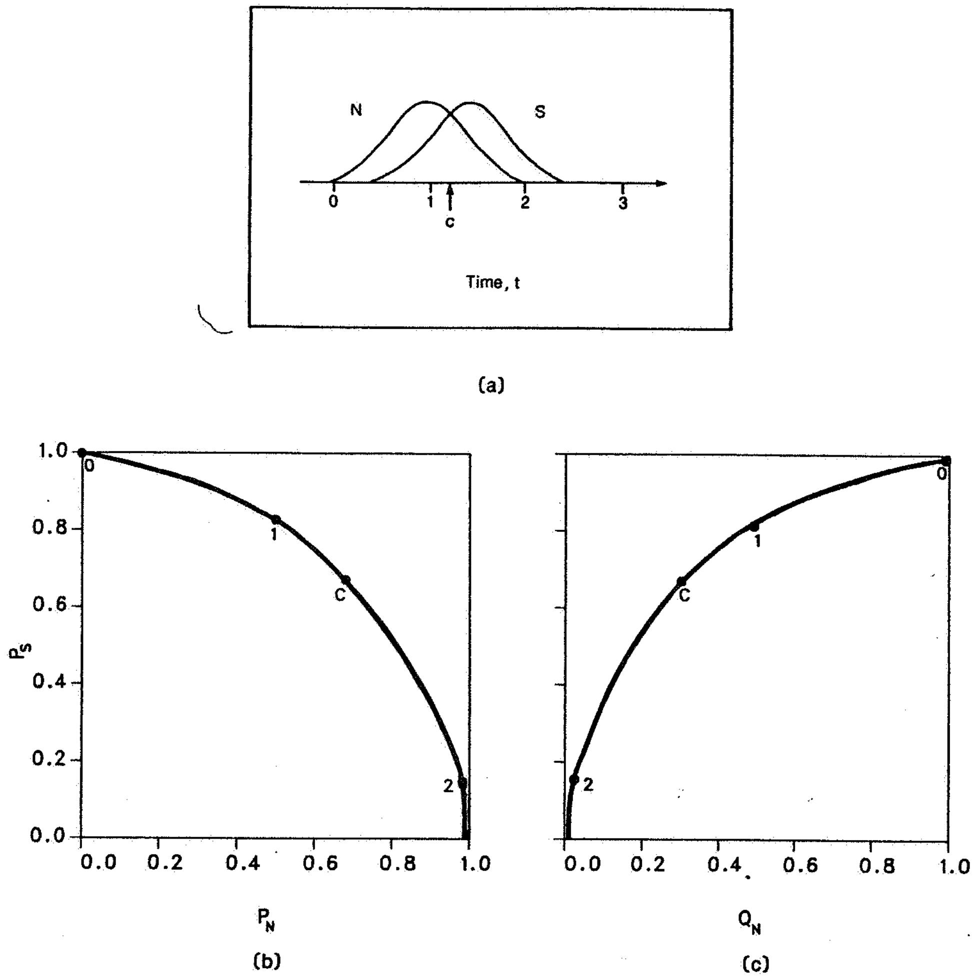



**Figure 2.3.** (a) A classroom example in which information densities as a function of time in two classes ($N,S$) are assumed to be normal distributions. The switching time for maximizing the total amount of acquired information is indicated by c. (b) The performance operating characteristic for a student who switches from class $N$ to class $S$ at various times is called an attendance operating characteristic (AOC) in this example. Representative switching times are indicated by points along the AOC. The axes represent the probability of correct responses ($P_N$ and $P_S$) on examinations in the two classes. (c) When the two distributions in panel (a) are assumed to represent the signal and noise distributions of a signal detection experiment, the POC obtained as the decision criterion is varied is called a Receiver Operating Characteristic (ROC). The abscissa in the ROC graph, $Q_N = 1 - P_N$, is $P(\text{"}S\text{"}|N)$, the probability of incorrectly saying "signal" when noise is presented; the ordinate, $P_S = P(\text{"}S\text{"}|S)$, is the probability of correctly saying "signal" when a signal is presented. Alternatively, panel (b) is the mirror image of panel (c) and therefore panel (b) can be interpreted either as a decision operating characteristic (DOC) for a signal detection experiment (the DOC is the mirror image of the ROC), or as a resource allocation operating characteristic for classroom attendance. (From G. Sperling, A unified theory of attention and signal detection, in R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

While psychological interpretation of the random variable as a sensory continuum is not essential to the application of SDT, intuitive interpretations are useful. When the two stimuli consist of a brief burst of noise, or noise to which a faint 1000-Hz tone has been added, then the random variable may be interpreted as the perceived amount of 1000 Hz on any trial. This perceived amount will vary from trial to trial but will tend to be slightly larger on signal trials than on noise trials. In visual experiments, the random variable may be interpreted as perceived intensity or contrast; in memory experiments, it might correspond to perceived familiarity. If in some unusual situation it should happen that the likelihood ratio [$lr(x) = p_S(x)/p_N(x)$] does not increase monotonically with $x$, then it is mathematically convenient (but psychologically incorrect) to consider $lr(x)$ (rather than $x$ directly) as the decision variable.

Figure 2.3(a), which we previously interpreted as representing information density, is now interpreted as representing the conditional distributions [$p_N(x); p_S(x)$] of the random variable on the sensory continuum $x$, where $x$ represents an amount of 1000 Hz, and so on. On any trial, $k$, the stimulus produces an effect $x_k$ described by a sample from the appropriate distribution ($p_S, p_N$) and the observer reports "signal" if $x_k > c$ and "noise" otherwise. The probabilities of a hit $P_S$ and of a false alarm $Q_N$ are given by

$$P_S = \int_c^\infty p_S(x)dx \quad \text{and} \quad Q_n = \int_c^\infty p_N(x)dx \; . \tag{5}$$

A graph of $P_S$ versus $Q_N$, shown in Figure 2.3(c), is called a *receiver operating characteristic.*

## 3.3. Decision Operating Characteristic

The graphic conventions for receiver operating characteristics (ROCs) produce mirror images of other performance operating characteristics. Compare Figures 2.3(b) and 2.3(c). In fact, the ROC uses a counterintuitive convention: good performance on signal trials (hits) is graphed against bad performance on noise trials (false alarms). The mirror image of an ROC plots correct performance on signal trials versus correct performance on noise trials, or $P_S$ versus $P_N = 1 - Q_N$. The mirror-image ROC is mathematically equivalent to the ROC but follows the more usual convention of graphing good performance up and to the right. Although SDT originally was applied to the discrimination of signals from noise, the formalism of SDT has been applied equally well to many other situations. For example, SDT applies well to discrimination experiments in which an observer's task is to discriminate two stimuli (e.g., tones of 1000 and 1001 Hz) as opposed to discriminating one stimulus from zero. The general case is discrimination (of which discrimination from zero, detection, is a subcase). The graph of $P_S(x)$ versus $P_N(x)$ as $x$ varies is appropriately called a *decision operating characteristic* (Sperling & Melchner, 1978b) or, more generally, a *performance operating characteristic* (Norman & Bobrow, 1975).

## 3.4. Operating Characteristics in Economics

Economic theory is useful in the attentional framework because it explicitly considers productivity given *limited resources.* Operating characteristics occur throughout economic theory; the relevant trade-off here is called the production possibility frontier (e.g., Samuelson, 1980). Figure 2.4 illustrates a hypothetical production possibility frontier for the strength of the military sector of an economic system measured in units of "swords" and agricultural output measured in "plowshares." The production possibility frontier results from a limited pool of resources, for example, citizens who could be either farmers or warriors, which can be allocated to either goal of the economy or shared between goals in various proportions.

Production possibility frontiers are generally concave toward the origin, as a result of the principle of symmetrically disposed diminishing returns. Diminishing returns mean, for example, that if the society were to increase the fraction of farmers from 0.95 to 1.0, it would increase the number of plowshares by less than when it increased the fraction of farmers from 0.0 to 0.05. Since diminishing returns for plowshares and swords occur on opposite ends of the production possibility frontier, decreasing the fraction of farmers from 1.0 to 0.95 would decrease plowshares by less than the increase in swords (see Figure 2.4(a)). It is assumed that the first persons to change occupations would be among the worst farmers, that is, persons who were relatively more efficient as warriors than as farmers. Similarly, if the fraction of warriors were to increase from 0.95 to 1.0, it would increase the number of swords only slightly, as the very last to join the army would be the least efficient soldiers relative to their efficiency as farmers.

The tendency to concavity toward the origin is a general property of performance operating characteristics that results from *unequal resources*—resources that are not equally interchangeable for all tasks. One way of expressing the inequality of resources is the performance resource function (Norman & Bobrow, 1975), a graph that describes the performance (e.g., agricultural production, plowshares) as a function of resources (e.g., farmers, number of acres, facilities, research) devoted to it. The discussion here is restricted to a single kind of resource, labor. Figure 2.4(b) shows the increase in agricultural production (plowshares) as a function of the number of agricultural laborers; it has a horizontal asymptote indicating that productivity ultimately is absolutely limited by factors other than labor. Figure 2.4(c) shows a graph $X - Y$ of the possible allocations of labor (number of workers) to the two competing sectors, swords and plowshares. The nonlinear performance resource function of Figure 2.4(b), which is the production possibilities frontier of Figure 2.4(a), is embodied in the utility function of Figure 2.4(c), yielding an equivalent description. The description in Figure 2.4(a) almost always is preferable to that in Figure 2.4(c).

In economics, production possibility frontiers, insofar as the two sorts of productivity can be measured, are considered to be objective descriptions. They are computed by the engineers and managers of the society. However, the *utility* of any joint combination of swords and plowshares depends, in principle, on the values and circumstances of the members of the society, although all may not contribute with equal weight. Equal utility contours are generally concave away from the origin because unbalanced combinations (strong military but no food or vast food supplied but no protection) are not as advantageous as balanced ones. The solution for the society, once the utility function is determined, is to find the point along the production possibility frontier where it touches the highest iso-utility contour. For smooth curves, the utility contour and the production possibilities frontier will be tangent to each other and of equal slope at the optimum point. Finding the optimum point along various trade-offs is at the heart of classical economic theory, and specialized branches of mathematics (such as linear programming) have been developed to deal with the problems of optimization. Section 3.5 explores the equivalences between trade-offs and optimization in signal detection, in concurrent attentional tasks, and in economic theory.

## 3.5. Optimization

**3.5.1. Signal Detection Theory.** Optimization in SDT traditionally involves the computation of the likelihood ratio $lr$. If $x$ represents values on the relevant internal decision axis, and the probability distributions over $x$ for Noise and Signal + noise conditions are $p_N(x)$ and $p_S(x)$, respectively, the likelihood ratio, given a particular value of $x$, is

$$lr(x) = p_S(x)/p_N(x) \ . \tag{6}$$

Figure 2.5 illustrates $p_S(x)$ and $p_N(x)$ as the equal-variance normal distributions of SDT. The decision criterion $c$ on the $x$-axis corresponds to a decision criterion of $\beta$ on the $lr$-axis, where the optimal $\beta$ is chosen according to prior probabilities and payoffs (Green & Swets, 1966). The decision operating characteristic (mirror-image receiver operating characteristic) at the upper right of Figure 2.5 represents the joint performance on $N$ trials and $S$ trials as $c$ is varied from $-\infty$ to $\infty$. The logarithm of the likelihood ratio is illustrated in Figure 2.5, center. The reason for illustrating log $lr$ rather than $lr$ itself is that log $lr$ is symmetric around $lr = 0$, which reflects the actual symmetry of treatment of $lr$ and $lr^{-1}$, and because log $lr$ occurs in many statistical treatments.

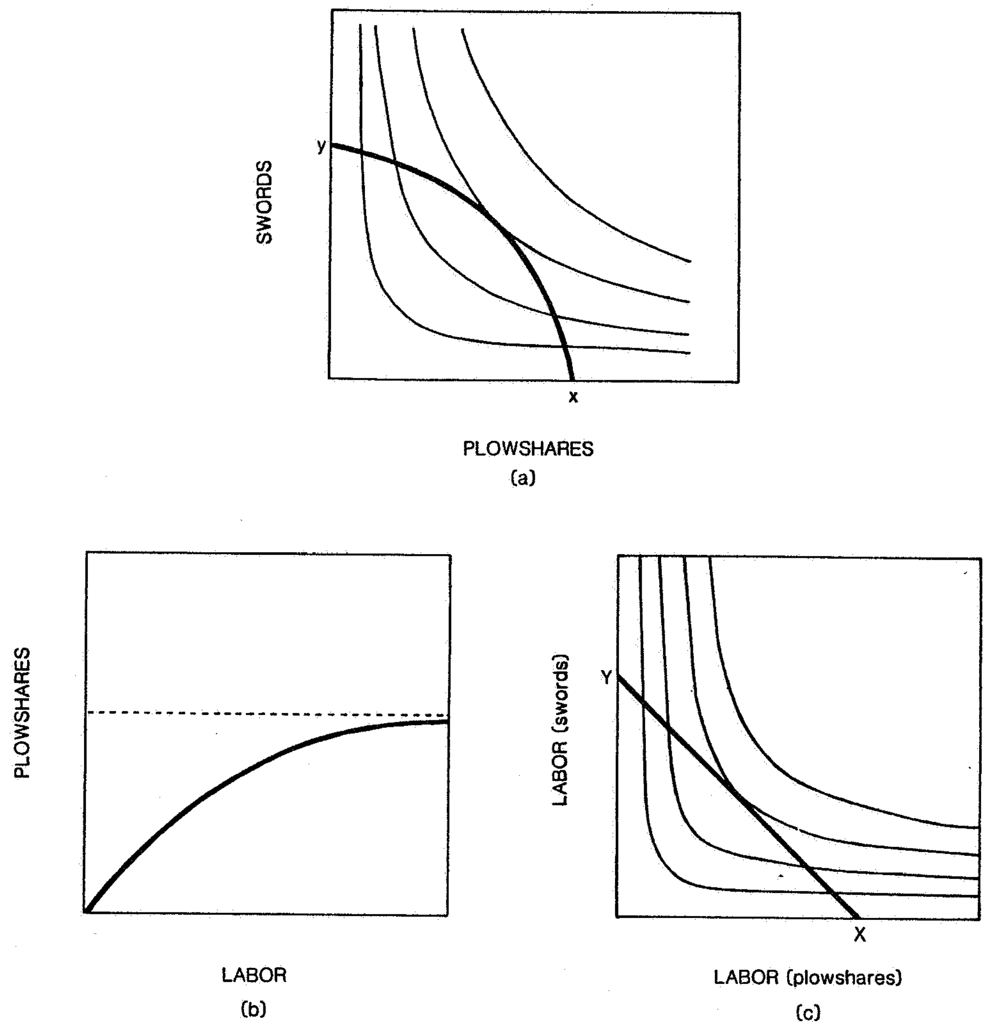



**Figure 2.4.** (a) A *production possibility frontier* (thick curve *xy*) for a primitive economy capable of producing swords or plowshares. The axes represent the number of plowshares and swords produced. Several iso-utility contours are shown as thin curves. The iso-utility contours are nearly flat on the extreme right because the utility of large changes in surplus plowshares can be compensated by small changes in scarce swords. The almost vertical ends of the iso-utility contours on the upper left represent symmetrically disproportionate marginal utilities for surplus swords. These iso-utility contours represent the principle of diminishing returns. (b) A resource-performance function illustrating agricultural production (plowshares) as a function of the amount of labor devoted to agriculture. (c) An alternative description of the swords-plowshares trade-off; a trade-off function (heavy line *XY*) showing the amount of labor devoted to swords versus labor devoted to plowshares. The thin lines represent iso-utility contours for these labor allocations; these iso-utility contours combine the nonlinear resource-performance function of panel (b). (From G. Sperling, A unified theory of attention and signal detection, in R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

The $d'$ statistic of SDT is the normalized distance between the mean of the $N$ and $S$ density functions; $d'$ summarizes the discriminability of $N$ from $S$. A more general statistic is the area under the receiver operating characteristic or the decision operating characteristic. The area under the decision operating characteristic, $A$, is:

$$A = P(x_S > x_N) \ . \tag{7}$$

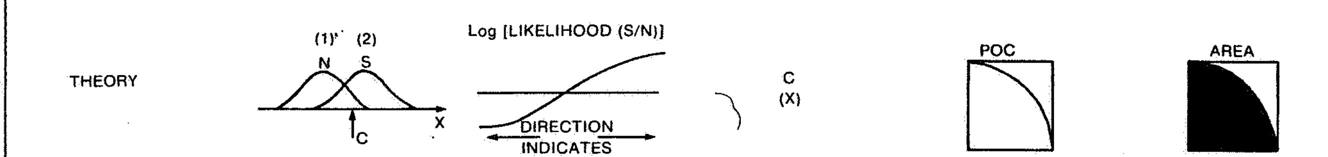
That is, $A$ is the probability that a random sample $x_S$ from the distribution $p_S$ exceeds a sample $x_N$ from $p_N$. According to a simple SDT model, $A$ is the probability of a correct choice in a two-alternative forced-choice (2AFC) task (Green & Swets, 1966). More generally, $A$ is a nonparametric measure of the amount by which the $S$ distribution dominates (is to the right of) the $N$ distribution. Statistics for $A$ are given in Bamber (1975). Signal Detection Theory (or decision theory) deals with compound tasks. The remaining cases considered in Figure 2.5 deal with concurrent tasks.

The following sections outline the reinterpretation of the critical concepts of SDT (the decision axis, probability densities, likelihood ratio, and area under the operating characteristic) when a similar optimization theory is applied to the classroom analogy, to economic production, and to attention. These analogies are developed in more detail in Sperling (1984).

**3.5.2. Attendance Theory.** In the classroom attendance example, two classes are offered during overlapping time periods and compete for the student's attendance. In this example, the

$x$-axis (decision axis of SDT) represents time, and the conditional distributions represent the usefulness of information offered at time $x$ for each of the classes ($N$ or $S$). The ratio of the two usefulness functions at any given time $x$ is analogous to the likelihood ratio of SDT. This ratio should be interpreted as the relative usefulness of spending the moment of time $x$ in class $S$ relative to class $N$. The classroom switching time in the classroom example is analogous to the criterion in SDT. The performance operating characteristic (POC) is generated by varying the classroom-switching time. Finally, the area under the POC has an interesting interpretation. If $x_S$ and $x_N$ are the times at which randomly sampled bits of information occur for class $S$ and class $N$, then the area $A$ under the POC is $A = P[x_S > x_N]$, the probability that a randomly sampled bit of information in class $S$ is offered later than a bit sampled from class $N$. This is a measure of the difference in times at which information in the two classes is actually offered.

**3.5.3. Economic Production Theory.** In economic theory, the SDT decision axis of "observations" (ordered in terms of their likelihood of indicating $N$ or $S$) is replaced by an ordering of resources (ordered according to their usefulness for the competing production goals of the economy). Let $f_1$ represent the usefulness (productivity) of a laborer as a farmer, and let $f_2$ represent his or her usefulness as a warrior. The ratio $x = f_2/f_1$ corresponds to the likelihood ratio of decision theory and can be used to order all the laborers or resources on a *usefulness ratio axis*. Those whose usefulness as farmers (relative to warriors) is greatest would be represented at the left side of the axis; those whose usefulness as warriors is greatest would be represented on the right side; $p(x)$ is the density function that indicates the fraction of laborers whose usefulness ratio is $x$. The usefulness ratio $f_2/f_1$, is sometimes called the *objective substitution rate*; in this example, it represents the rate at which swords can be substituted for or converted into plowshares.




| | | | | | |
|---|---|---|---|---|---|
| SIGNAL DETECTION SDT | Density function for conditional probability that sensory observation = x, given N, given S | Noise  Signal | Criterion (sensory continuum) | ROC Receiver Operating Characteristic | P ($X_S > X_N$) |
| DECISION | Density function for conditional probability that sensory observation x, given stimulus 1, given stimulus 2 | Stimulus 1  Stimulus 2 | Criterion (decision continuum) | DOC Decision Operating Characteristic | P ($X_2 > X_1$) |
| ATTENDANCE | Density of classroom information, given class 1, given class 2 | Usefulness for class 1  Usefulness for class 2 | Classroom switching (time) | AOC Attendance Operating Characteristic | P (info in class 2 is offered after info in class 1) |
| ATTENTION | Net productivity of resources for tasks 1,2 | Usefulness for task 1  Usefulness for task 2 | Allocation decision (resources) | AOC Attention Operating Characteristic | P (a task 2 resource is more useful for task 2 than a task 1 resource is) |
| ECONOMIC PRODUCTION | Net productivity of resources for tasks 1, 2 | Usefulness for producing product 1  Usefulness for producing product 2 (Objective substitution rate) | Allocation decision (resources) | PF Production Possibilities Frontier | P (a resource used to produce 2 is more efficient in producing 2 than is a resource used to produce 1) |

**Figure 2.5.** A summary of the isomorphisms between five theories: signal detection, decision (compound tasks), attendance, attention (concurrent tasks), and economic production. Density functions are indicated by $N(x)$, $S(x)$. The log likelihood ratio is $\ln[S(x)/N(x)]$; it is a *log usefulness* ratio in the resource theories. The dimensional interpretation of the variable x is indicated by the term in parentheses; c indicates the decision criterion. The performance operating characteristic (POC) is the curve indicated in the graph. The shaded area in the rightmost graph represents the area $A$ under the POC; $A = P(X_2 > X_1)$ represents the probability of the event that a randomly chosen sample from distribution 2 is greater than a randomly chosen sample from distribution 1. The interpretation of each of these quantities in each theory is indicated in the figure. See the text for details. (From G. Sperling, A unified theory of attention and signal detection," in R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

Productivity functions analogous to the density functions $p_N(x)$ and $p_S(x)$ of SDT are derived as follows. The fraction of laborers with usefulness ratio between $x$ and $x + dx$ is $p(x)dx$. The average usefulness for agricultural production of these laborers is $f_1(x)$. The net productivity of this group is $p_1(x)dx = f_1(x)p(x)dx$; $p_1(x)$ is a density function analogous to $p_N(x)$ of SDT. (In order for $p_1(x)$ to represent a density function, $f_1(x)$ has to be appropriately scaled, and there are other technical restrictions that are not central to the discussion here.) The density function $P_1(x)$ represents the net productivity as farmers, $p_1(x)dx$, of all the laborers whose objective substitution rate (or usefulness ratio) is between $x$ and $x + dx$. The function $p_2(x)$, analogously defined, represents the productivity as warriors of this same group. Productivity is measured in normalized units, that is, in terms of the fraction of the maximum capacity for the specified task.

The decision criterion $c$ corresponds to the decision to assign all laborers with usefulness ratio less than $c$ to farming and the remainder to fighting. The decision operating characteristic of SDT, generated as $c$ varies from $-\infty$ to $+\infty$, corresponds to the production possibilities frontier of economic theory, similarly generated.

Finally, the area $A$ under the production possibilities frontier has a similar interpretation to the areas under the decision operating characteristic and the performance operating characteristic: it represents the probability that a randomly chosen sword-resource unit will be more useful for sword production than a randomly chosen plowshare-resource one would be for sword production. $A$ is a measure of the extent to which skills or facilities (e.g., for farming, fighting) are segregated into different people or facilities (nonsubstitutable) as opposed to coexisting in the same person or facility (substitutable). A similar analysis in terms of economic consumption by laboratory animals (where there is only one resource, e.g., time, to allocate to separate appetites) is treated in detail by Rachlin, Battalio, Kagel, and Green (1981) and Sperling (1984). Rachlin, Green, Kagel, and Battalio (1976), Rachlin, Kagel, and Battalio (1980), and Rachlin and Burkhard (1978) give empirical examples of substitutability and interference.

## 3.6. Attention Theory

**3.6.1. Concurrent Tasks.** Attentional allocation is closest in interpretation to economic production theory. The allocation of mental resources is assumed to determine the quality of performance of several concurrent cognitive tasks, just as the allocation of economic resources determines the extent to which competing manufacturing goals are achieved. The application of economic theory to attention was proposed by Navon and Gopher (1979). The direct application of the economic analogy is appropriate only for *concurrent* tasks; in compound tasks, the effects of stimulus uncertainty must first be removed.

The critical aspect of attention theory is the interpretation of the decision axis as an ordering of resources—in the case of attention, *mental processing resources*. The units of mental resources are defined analogously to those of production resources. Allocation of any single unit of Task 1 resources to Task 1 will accomplish, say, 1% of the maximum achievable performance for Task 1. Of course, those particular resources may be more or less effective for Task 2. The mental resources whose usefulness ratio $x$ (usefulness for Task 2 divided by usefulness for Task 1) is lowest are represented at the extreme left of the resource axis (Figure 2.5). Thus the resource axis is directly analogous to a likelihood decision axis of SDT. The conditional density function $p_1(x)$ represents the usefulness or productivity for Task 1 of resources as a function of their *usefulness ratio* $x$; $p_2(x)$ represents the usefulness of resources for Task 2. (More precisely, the area $p_i(x)dx$ under the density function represents the aggregate usefulness of resources whose usefulness ratio lies between $x$ and $x + dx$.) The functions $p_1(x)$, $p_2(x)$ are both normalized relative to the maximum achievable performances. The decision by the subject to allocate mental resources with usefulness ratio less than $c$ to Task 1 and the remainder to Task 2 is analogous to the decision criterion $c$ in SDT. The attention operating characteristic is traced out as $c$ is varied over its range by attentional manipulations (e.g., instructions to attend to Task 1 versus Task 2). The area under the attention operating characteristic represents the probability that a resource unit, chosen at random from all those useful for Task 2, really is more useful for Task 2 than a randomly chosen Task 1 resource unit would have been. It is a nonparametric measure of the extent to which distinct (nonsubstitutable), as opposed to interchangeable (substitutable), resources are involved in performing the two tasks.

**3.6.2. Compound Tasks.** Attentional manipulations can be interpreted as controlling resource allocation only in concurrent tasks. In compound tasks, because of the effects of stimulus uncertainty, the attentional manipulation must first be viewed as a decision manipulation subject to decision uncertainty (as in signal detection or decision theory, Sperling, 1984). More qualitative arguments along these lines can be found in Duncan (1980). Stimulus uncertainty in compound tasks is treated extensively in Section 5. If, after stimulus uncertainty has been accounted for, there is a residual effect of attention in a compound task, then obviously resource analysis would be appropriate for this residual effect.

## 3.7. Single and Multiple Resources

**3.7.1. Undifferentiated versus Differentiated Attention.** Early in the development of attention theory, concurrent task performance was interpreted with respect to an undifferentiated capacity hypothesis (Kahneman, 1973; Moray, 1969). According to this hypothesis, interference between tasks occurs when the total demands exceed the pool of attentional (processing) capacity. This pool is undifferentiated with respect to the specific operations of the particular tasks. Alternatively, Treisman (1969) suggested a structural attention model, where tasks interfere only to the degree that they call upon the same processing "subsystems" or "analyzers." The formal attention theory described here is a differentiated model; the model represents an ordering of processing resources according to the usefulness ratio for two competing tasks. Relative usefulness is a continuous concept and more general than the structural model of Treisman. The undifferentiated capacity model corresponds to the special case in which all resources have precisely the same usefulness ratio for all the competing tasks.

In the language of the classroom analogy, an undifferentiated capacity model could be conceived as follows. The limitation in resources is a maximum number of hours $m$ to be spent in class. Since the capacity is undifferentiated, a subject would be free to arrange two classes to occur in the optimal arrangement within the $m$ hours (i.e., the $x$-axis is arbitrary), and interference would occur only if even the optimal arrangement involved overlapping class times.

The undifferentiated capacity model makes a strong prediction about the relation to each other of the attention operating characteristics (AOCs) produced by the three binary combinations of three tasks. This situation is outlined in Figure 2.6, which shows some hypothetical arrangements of classes A, B, and C. Under the undifferentiated capacity model, once the AOCs for concurrent tasks (A,B) and for (A,C) have been measured, the AOC for the concurrent tasks (B,C) is completely specified. See Figure 2.6 (a), (b), (c), (d), (g), and (h).

No such prediction is made by the differentiated processing resource model specified in this chapter. Sample configurations of classes for the differentiated processing resource model are shown in Figure 2.6. Two extremes are shown. In case (e)-(f), the classes B and C both partially interfere with A, but they

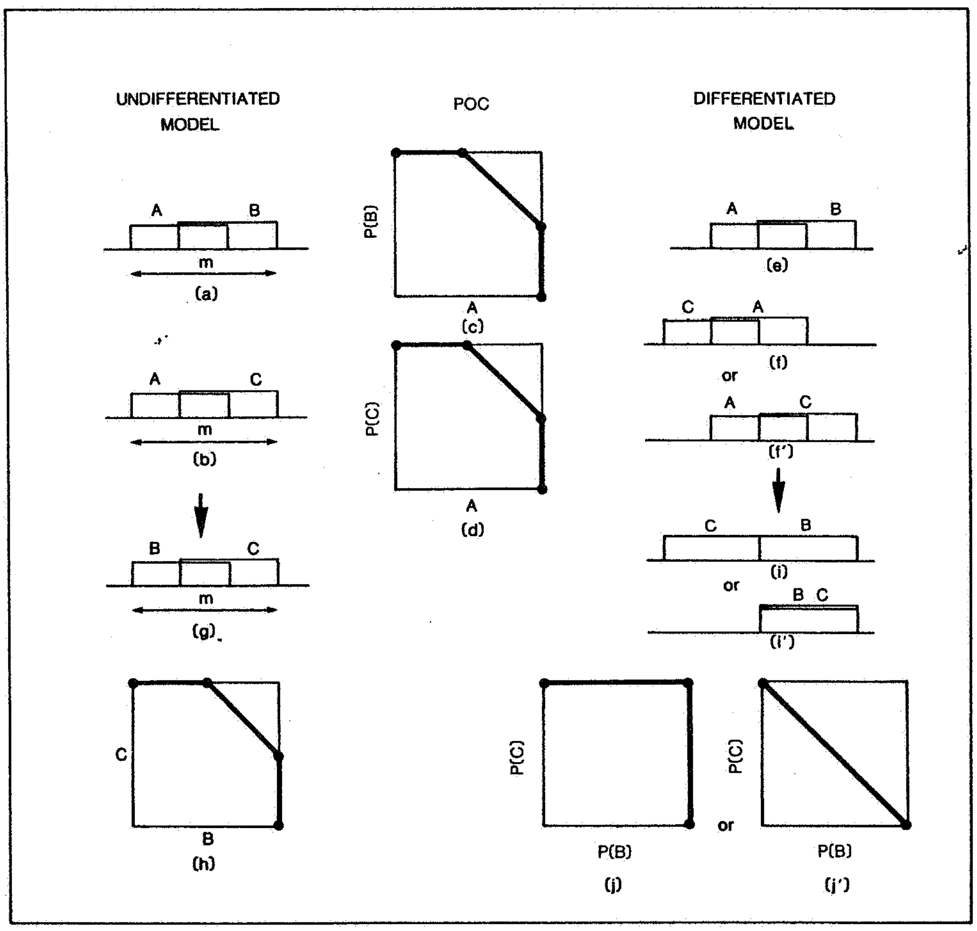



**Figure 2.6.** The undifferentiated and differentiated attentional models. The *undifferentiated capacity model* assumes that all processing resources are interchangeable, but that there is a limitation in the total amount of resources. These assumptions are embodied in an example of tutorial instruction in which *m* is the capacity limit—the maximum number of student hours available for attending tutorial instruction. Tutorial instruction can be arranged freely (up to a maximum with each tutor) within the *m*-hour limit. Panels (a) and (b) indicate various tutorial schedules a student can arrange involving tutors (*A*, *B*) and (*A*, *C*). The performance operating characteristics (POCs) resulting from varying the proportion of *A* in (*A*, *B*) and (*A*, *C*) are shown in panels (c) and (d). When task combinations (*A*, *B*) and (*A*, *C*) are known to produce the POCs in (c) and (d), then the undifferentiated capacity model first specifies that the tutors (*B*, *C*) can be scheduled, for example, as shown in panel (g), and second, uniquely specifies that the resulting POC from *any* reasonable schedule is that shown in panel (h). The *differentiated model* assumes that resources can be differentially useful for various classes. Possible classroom information density functions under the differentiated model that would produce the AOCs in panels (c) and (d) are shown in panels (e) and (f or f'), respectively. If panels (e) and (f) describe the overlap of (*A*, *B*) and (*A*, *C*), then the overlap of (*B*, *C*) is shown in panel (i) and the corresponding AOC is shown in panel (j). If, instead, panels (e) and (f') describe the overlap of (*A*, *B*) and (*A*, *C*), then the overlap of (*B*, *C*) and the corresponding AOC appear in panels (i') and (j'). The same (*A*, *B*) and (*A*, *C*) AOCs yield radically different AOCs for the (*B*, *C*) class combination. Any AOC between the extremes, shown as panels (j) and (j'), is possible under the differentiated model. However, if classes (*A*, *B*) and/or classes (*A*, *C*) overlap more than the examples in panels (e) and (f), the differentiated model constrains the possible (*B*, *C*) AOCs, in a manner analogous to the constraints on correlations between three random variables (Kendall, 1970). In the limit, if (*A*, *B*) overlap completely and (*A*, *C*) overlap completely, then (*B*, *C*) must overlap completely in both the undifferentiated and the differentiated models.

require different resources, and therefore B and C do not interfere with each other. In case (e)-(f′), B and C partially interfere with A by virtue of requiring identical resources, and therefore they interfere totally with each other.

The usefulness ratio is inherently defined for a particular task combination (A,B). Another task combination (A,C) need not result in the same ordering of processing resources. Both the cases considered for the differentiated model in Figure 2.6 actually assume some similarity in the usefulness-ratio axis for the (A,B) and (A,C) combinations. However, in the general case each task combination reflects the task distribution on a potentially *unique* usefulness-ratio axis. Therefore, knowing the AOC for (A,B) and (A,C) specifies only a limited constraint about the task combination (B,C), which could fall somewhere between the two special cases shown in Figure 2.6. In the special case where classes (A,B) overlap completely and classes (A,C) overlap completely, both the differentiated and the undifferentiated models make the same clear prediction: (B,C) also must overlap completely.

**3.7.2 Nature of Mental Resources.** What are mental resources? There are two approaches to this question. The first is that it is not necessary to know what mental resources are. Mental resources may have the status of a random variable much like the decision variable of SDT. All the power and prediction of SDT work whether or not the psychological (mental) dimensions of the decision variable are precisely known. All the power of optimization theory is available to predict and describe performance in concurrent tasks even when it is not known precisely where these tasks conflict. However, cognitive psychologists have a special interest in learning precisely what particular mental resources are involved in cognitive functions.

With respect to particular mental resources, the critical resources for which there is competition vary with the task. In the partial-report or concurrent whole-report tasks (Section 7.4.1), the critical resource is short-term memory, it has a limited capacity, and that capacity is allocated to items from one stimulus row or the other according to the task demand. This memory resource seems to be quite interchangeable.

In search tasks (Section 5), the critical resource probably is a processing resource involved in making comparisons. A stimulus item at one location in the visual field *can* be compared to a memory representation of a target at the same time that another item in another part of the field is being compared to a representation of another target. However, the extent to which such comparisons draw on different resources (and therefore can occur simultaneously) and the extent to which they draw from a common pool of resources (and therefore must be made serially) depends on many factors, among the most important of which is the familiarity of the target—the extent to which special resources have been developed for particular targets (Shiffrin & Schneider, 1977). The issue of serial versus parallel processing is a fascinating research problem; for an example of recent theorizing involving queuing theory, see Fisher (1982). Section 8.6 examines a powerful method of determining whether resources from a common pool can be evenly shared by two tasks or whether they are switched in all-or-none fashion from one task to the other on different trials.

## 3.8. Iso-utility Contours

Iso-utility contours are a powerful heuristic device for studying optimization. They have long been used in economic theory to investigate which of a number of alternative procedures or parameters produces the maximum utility or most preferred outcome. Navon and Gopher (1979) introduced iso-utility contours into the study of attention. They were introduced here in the classroom example of Section 2. Their use in signal detection experiments remains to be specified.

On each trial of a signal detection task, either a signal or a noise stimulus is presented and the subject responds either "*S*" or "*N*." The experimenter explicitly or implicitly assigns a utility to each of the four possible outcomes of a trial. This is commonly called the *payoff matrix*, as shown in Table 2.1. Let the fraction of signal trials be $\alpha$; then the fraction of noise trials is $1 - \alpha$. The subject's probability of saying "*S*" given *S* (i.e., detecting the signal when it is presented) denotes $P(\text{"}S\text{"}|S)$; $P(\text{"}N\text{"}|N)$ is analogously defined. Given the payoff matrix in Table 2.1, the expected utility *EU* of a trial is

$$EU = \alpha[dP(\text{"}S\text{"}|S) + cP(\text{"}N\text{"}|S)] + (1 - \alpha)[bP(\text{"}S\text{"}|N) + aP(\text{"}N\text{"}|N)] \ . \tag{8}$$

Equation (8) gives the expected utility for every possible performance level, $P(\text{"}S\text{"}|S)$ and $P(\text{"}N\text{"}|N)$ for particular values of $a$, $b$, $c$, and $d$. The limit on achievable performance levels is described by the discrimination operating characteristic, which is a graph of $P(\text{"}S\text{"}|S)$, $P(\text{"}N\text{"}|N)$ pairs obtained as some nonstimulus parameter of the experiment is varied. (The operating characteristic is the "limit of performance" since the subject could follow a nonoptimal decision strategy.)

To illustrate the effect of $\alpha$ on performance, we choose a particular payoff matrix; for example, wrong responses $b$ and $c$ earn zero ($b = c = 0$) and correct responses $a$ and $d$ earn 1 dollar per trial ($a = d = 1$). Figure 2.7(a) illustrates iso-utility contours for $\alpha = 0.25$, and Figure 2.7(b) illustrates iso-utility contours for $\alpha = 0.75$. The parameter on the contours is the expected utility. Expected utility as defined by the payoff matrix and Eq. (8) is computable for all values of $P(\text{"}S\text{"}|S)$, $P(\text{"}N\text{"}|N)$, not just achievable values. The iso-utility contours are straight lines with slope $M$, where $M$ is

$$M = \left(\frac{b - a}{d - c}\right)\left(\frac{1 - \alpha}{\alpha}\right) . \tag{9}$$

**Table 2.1. Payoff Matrix for the Four Possible Outcomes of a Trial in a Signal Detection Experiment**

| | | Response: "Noise" | Response: "Signal" |
|---|---|---|---|
| Stimulus | N | $a$<br>Correct rejection | $b$<br>False alarm |
| | S | $c$<br>Miss | $d$<br>Hit (correct detection) |

*Note*: $a$, $b$, $c$, $d$ are real numbers that represent the payoffs (utilities).

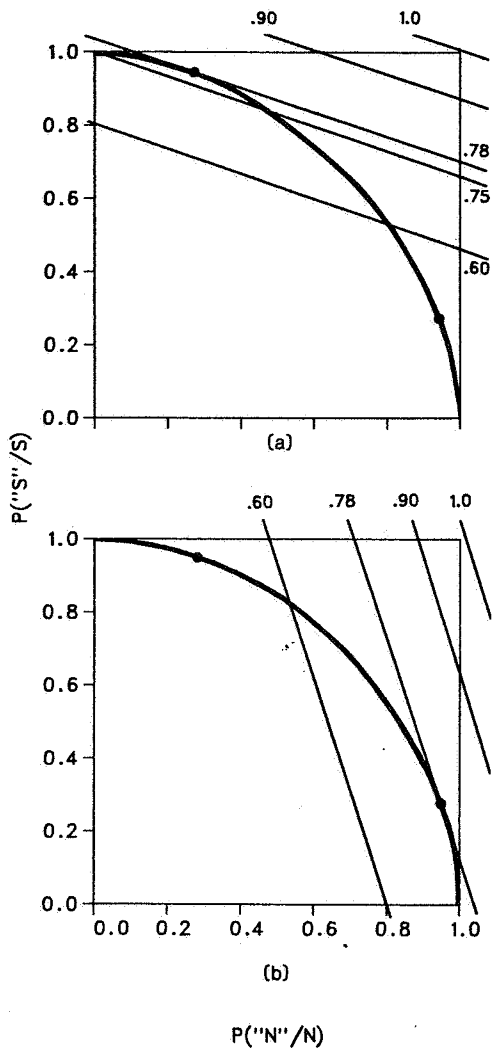



**Figure 2.7.** The influence of a priori signal probability on performance. Panels (a) and (b) show the same discrimination operating characteristic (DOC) for a signal detection task. The abscissa is the probability of a correct response given a noise stimulus, and the ordinate is the probability of a correct response given a signal stimulus. The DOC shown assumes equal-variance Normal distributions and a $d'$ of 1. In (a), the a priori probability of a signal is 0.75, the utility is 1 for correct responses and 0 for incorrect responses. The iso-utility contours show the expected utility (payoffs). Panel (b) shows the iso-utility contours when the a priori probability of a signal is 0.25. (From G. Sperling, A unified theory of attention and signal detection, in R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

If the outcomes of a trial are symmetrical with respect to $S$ and $N$ for both errors and correct responses, then the iso-utility slope is simply the ratio of the two a priori stimulus probabilities, $-(1-\alpha)/\alpha$. In the two examples of Figure 2.7, the slopes are $-3$ and $-1/3$.

A typical decision operation characteristic based on the assumption of equal-variance normal distributions for $N$, and $S + N$ also, is illustrated in both panels of Figure 2.7. From the graph, it is obvious that the criterion should be adjusted quite differently to achieve the optimal performance with $\alpha = 0.25$ than with $\alpha = 0.75$. The expected utility of each strategy (each criterion value) can also be estimated from the graph. The optimal strategy has an expected utility of 0.78 dollars per trial. This is only marginally better than 0.75, the utility that could be achieved by simply naming the a priori more probable stimulus on each trial without actually observing the stimulus presentation. For normally distributed signal and noise, it always pays to observe the stimulus because the likelihood ratio varies between 0 and $\infty$. But there are many distributions, such as the logistic distribution, for which the likelihood ratio is bounded. For stimuli characterized by distributions like the logistic, it can be better *not* to observe the stimulus when the a priori probabilities are very asymmetrical, but merely to use the a priori information.

The value of a priori information is the expected value of a trial with this information minus the value of a trial without it. In the example of Figure 2.7, suppose that an observer has no information about the a priori stimulus probabilities and therefore sets the criterion symmetrically (at a likelihood ratio equal to 1.0). The expected probability of a correct response would be 0.69 which, in this example, is also the number of utility units (dollars) the observer would expect to earn on each trial. Note that 0.69 is the highest achievable expected probability of a correct response with equally probable stimuli or with unequally probable stimuli when the probability is unknown. The a priori information that one stimulus is 3 times more probable than the other enables the observer to achieve an expected probability of a correct response of 0.75 without even observing the stimulus and 0.78 if choosing to actually observe it. The a priori information alone is thus worth more (0.75) than the opportunity of viewing the stimulus without a priori information (0.69). Finally, it is obvious that good detection of signal stimuli $P(\text{"}S\text{"}|S) \approx 1$ can be profitably exchanged for good detection of noise stimuli $P(\text{"}N\text{"}|N) \approx 1$ when there are more noise than signal stimuli, and vice versa.

All these properties and relations of variables in signal detection are, of course, derivable algebraically, and they are well known. The concept of utility is central to SDT (Swets, Tanner, & Birdsall, 1961). Furthermore, receiver operating characteristics have been graphed with various performance criteria (Swets, 1973) but not generally with iso-utility contours. (The sole exception is Metz, Starr, Lusted, & Rossman, 1975, Figure 5, p. 420). The aim here is to illustrate the properties of SDT and utility theory in a new way so that previously unobserved similarities between all the various kinds of situations (detection, discimination, attendance, attention, economics, etc.) are made explicit.

## 4. COMPOUND TASKS: UNCERTAINTY IN DECISION

A compound combination of several subtasks is defined as a task on which the stimulus on any trial is drawn randomly from the set of stimuli of any of the component tasks. Thus uncertainty is introduced in any compound combination. It was asserted earlier that decrements in performance in *compound* tasks can be interpreted as attentional decrements only after discounting the effects of decision uncertainty in an ideal observer. A computational model of uncertainty effects is required for interpretation. This section reviews the uncertainty loss of an ideal observer as it is treated in SDT (Egan, 1975; Swets, 1964). It then examines uncertainty and attentional effects as they apply in several kinds of detection experiments. The concept of uncertainty is extended to classification experiments, in general, and to experiments in visual search.

## 4.1. Examples of Compound Detection Experiments

**4.1.1. Auditory Detection Example.** In an auditory signal detection experiment, a 500-Hz tonal signal plus noise or noise alone occurs for 0.5 sec on each trial. The stimuli are $S + N$, $N$. The responses are "$S$," "$N$." This is component Task 1. In component Task 2, the stimuli are 810 Hz + $N$, $N$; in component Task 3, 1320 Hz + $N$, $N$. In the compound task, the stimuli are 500, 810, or 1320 Hz, each plus noise or noise alone. The responses are "$S$," "$N$." That is, the stimulus set is the union of the stimuli of the three component tasks; this is the special case, Condition (CP-4), where the response sets of the compound task and each component task are equal. A more elaborate version of this compound task would require the subject to make the response "$S$" or "$N$" and then to "recognize" the signal, that is, say "low," "medium," or "high" pitch.

Uncertainty decrements are almost always observed for frequency mixtures. The question is whether these decrements are "attention" decrements or the result of decision uncertainty. Auditory detection experiments with multiple alternative signals and single-band versus multiple-band interpretations of these experiments are extensively reviewed in Swets (1984). The analogous visual paradigms are considered in this section.

**4.1.2. Visual Detection Example.** One possible uncertainty experiment in the visual domain compares "$S$"/"$N$" detection of sine-wave gratings of various frequencies in "alone" blocks (the component tasks) with "$S$"/"$N$" detection in intermixed blocks (the compound experiment). This is analogous to the auditory frequency uncertainty experiment described in Section 4.1.1. In the visual domain, the compound experiments have shown decrements for intermixing of widely separated spatial frequencies and for intermixing of separated spatial positions but not for intermixing of contrast levels (Davis, Kramer, & Graham, 1983).

**4.1.3. Location Experiments.** This section considers in more detail the case of attending to several spatial locations. Wundt, in his introductory psychology text (1912), described a self-experiment for observing the spatial distribution of attention. The reader was instructed to fixate his or her eyes on a mark at the center of an array of letters and to direct his or her attention to a letter off to the side. Wundt asserted that the letters around the attended peripheral location appeared more vivid than those elsewhere in the array. Unfortunately, Wundt's dependent variable (judged vividness) is problematical; certainly it is not an objective measure of a *performance* that is affected by selective attention.

***4.1.3.1. Simple Yes-No Detection.*** One of the first serious experimental attempts to measure the spread of visual attention to several locations was by Mertens (1956). He required his observers to maintain fixation faithfully on a central fixation mark. He then presented them with very weak flashes of light to be detected. When they detected a flash, they indicated so by pressing a button. In some blocks of trials, the flashes could occur at any of four locations around fixation (northwest, southwest, southeast, or northeast), in others, only one (say, southwest). See Figures 2.8(a) and 2.8(b). Mertens's observers seemed to have slightly lower detection thresholds at an unknown one of four locations than at one predetermined location. He concluded that it was more effective for the subject to allow attention to spread out over four locations than "to stress himself continually not to look in the direction of attention" (p. 1070). Unfortunately, Mertens was unaware of the rudiments of SDT, so

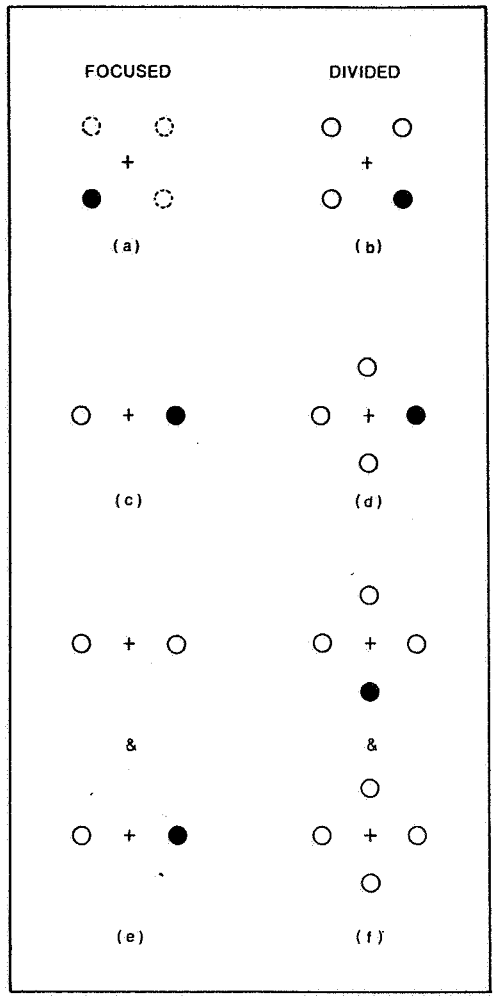



**Figure 2.8.** Stimulus configurations for visual detection experiments that compare focused with divided attention. Each panel illustrates a particular potential stimulus. A plus sign indicates the visual fixation point; a filled circle indicates a target; an open circle indicates a noise (nontarget) stimulus. (a) *Yes-No detection judgments*: Mertens's stimuli for focused attention. In the focused condition, targets could occur only in the particular location selected by the subject; the dashed open circles indicate that the other locations were displayed nevertheless. The subject's task is to say whether the target occurred on that trial. (b) In Mertens's divided attention condition, the target could occur at an unknown one-of-four possible locations; it is shown here in the southeast location. (c, d) *Forced-location judgments*. The target occurs on every trial in one-of-two locations in panel (c) or in one-of-four in panel (d), and the subject must say *where* it occurred. (e, f) *Two-interval forced-choice procedure*. Two temporal intervals (separated by & in panels) within which the target can occur are defined for the subject, but the target occurs only in one. The subject must identify the interval. (From G. Sperling, A unified theory of attention and signal detection, in R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

there were flaws in his procedure, such as inadequate treatment of false alarms. It is possible that the observer altered his criterion ($c$ in Figure 2.3(a)) between conditions but that $d'$ (the separation between $p_S(x)$ and $p_N(x)$) was not affected (or was even oppositely affected) by the attentional manipulation. Mertens's strange result was replicated once by Schuckman (1963) in an experiment with the same difficulties as Mertens's and by Howarth and Lowe (1966), who found no effect of any kind of uncertainty, not of stimulus location, size, or time of occurrence. Since then the opposite result has been obtained.

We now consider some contemporary approaches to spatial uncertainty in detection.

***4.1.3.2. Forced Location Judgment.*** Consider first the following Gedanken experiment. In some sessions a flash can occur in either of two locations (east or west); in other sessions it occurs in any of four locations. See Figures 2.8(c) and 2.8(d). The accuracy of naming the target location is measured and found to be higher in two-location than in four-location sessions. Unfortunately, it requires a theory of guessing to estimate the attention effect, since visual detection accuracy is higher in two-location than in four-location stimuli even when the observer's eyes are shut and, for low-intensity stimuli, the guessing effect is predominant. To obviate guessing analysis, a two-alternative forced-choice paradigm may be used (Cohn & Lasley, 1974). All trials are composed of two intervals, and a target always occurs in one or the other of the intervals. In some sessions there are two possible locations in which the target may occur; in others there are four possible locations for the target, as seen in Figures 2.8(e) and 2.8(f). Subjects correctly identify the interval containing the target more frequently in the two-location than in the four-location trials. Since chance guessing is the same in both kinds of trials, this result appears to demonstrate an attentional loss in attempting to monitor four locations. Unfortunately this conclusion is premature, as shown in Section 4.2.

## 4.2. Signal Detection Theory and the Ideal Observer

In all these examples, the question that must be answered is whether the performance loss in the compound experiment compared to performance in its components is due to a *limitation of attention* or to *uncertainty in the decision process*. Signal Detection Theory predicts a decrement in performance whenever the subject is confronted with a larger number of independent noise samples.

**4.2.1. Uncertainty in Location.** Consider a detection task in which the target can occur at one of a number of possible locations. Each location $i$ that the subject must monitor is assumed to produce a sample $n_i$ from a noise distribution. At the target location $t$, the signal is added to noise to produce $n_t + s$. The decision rule for the ideal detector (assuming that the target occurs with equal probability at all locations) is to choose the location with the largest sample. If $n_t + s$ is greater than $n_i, i \neq t$, then a correct location detection would occur. If, however, $n_i > (n_t + s)$ for some $i \neq t$, a false detection would occur. The probability of false detections will increase with the number of locations, even when the quality of information about each location remains the same. Quantitative predictions, of course, depend on the assumed shape of the noise distribution.

Intuitively, an ideal detector makes a mistake when the noise sample at some nontarget location exceeds the sample of signal plus noise at the target location. If the decision rule chooses the location with the largest sample, it is necessary to consider only the largest noise sample. The two-interval (Cohn & Lasley, 1974) paradigm described previously compares the maximum of three noise samples (in the two-interval–two-location case, Figure 2.8(e)), with the maximum of seven noise samples (in the two-interval–four-location case, Figure 2.8(f)). It is known from the general properties of order statistics that the expected value of the maximum of $n$ independent, identically distributed random variables increases as the $n$ increases, and hence the chance of an error in detection increases. In the two-interval paradigms, a response based on a noise sample may be either correct or incorrect, but a response based on the signal will necessarily be correct. The more locations monitored, the larger the number of noise samples, and the larger the number of errors. One advantage of concentrating on the largest noise sample (subject to some technical restrictions; see Gumbel, 1958) is that while the distribution of the noise random variable may be unknown, in the limit as $n \rightarrow \infty$ there are only three possible distributions of maxima (Galambos, 1978; Gumbel, 1958). When the number of monitored locations is large, the distribution of the maximum noise sample may be better known than that of the individual samples.

**4.2.2. Distribution-Free Bounds on Location Uncertainty.** A special case of the location uncertainty problem has been considered by M. L. Shaw (1980). She derived a distribution-free prediction for the maximum decision (stimulus uncertainty) loss for experiments involving two or $M$ locations in which the subject's task is to name the target's *location*. This computation is valid only when a test of two-location identification indicates that the subject is monitoring both locations to some degree. Let $P_2$ and $P_M$, respectively, represent the probabilities of a correct location judgment in 2- and $M$-location experiments, respectively. If the data from a 2- versus 4-location experiment fall below the distribution-free boundary shown in Figure 2.9 or more generally if $P_M < (P_2)^{M-1}$, then the explanation must involve more than decision uncertainty. Figure 2.9 also illustrates predictions of the stimulus uncertainty loss for the 2- and 4-location cases assuming exponential or normally distributed noise. The exponential and normal computations illustrate the large range of stimulus uncertainty loss possible in compound tasks and the corresponding problems in interpreting data. Although data falling below the distribution-free boundary indicate a performance loss that cannot be due only to stimulus uncertainty (and therefore could be attentional), this is a very weak test. Data falling anywhere between the distribution-free boundary and no loss (Figure 2.9) are ambiguous; they exceed the normative stimulus uncertainty loss under some distributional assumptions but not others.

**4.2.3. Uncertainty in Detection of Independent Signals.** One approach to the problem of uncertainty in *detection* of one of $M$ *independent* (orthogonal) signals is to make the computations based on the assumptions of equal-variance Normal noise distributions (Nolte & Jaarsma, 1967). Suppose it is known that, on signal trials, only one of $M$ independent signals (plus noise) will be presented, and the problem is to discriminate signal trials from noise trials. Optimally, the observer will construct a likelihood ratio $lr$ that combines the information from each of the $M$ independent samples, or "channels." For noise trials, $lr$ reflects $M$ independent and identically distributed noise samples; for signal trials, $M - 1$ noise samples and the signal-plus-noise sample. Under the assumption of equal-variance Normally distributed noise and signal plus noise, Nolte and Jaarsma (1967) derived receiver operating characteristics for several values of $d'_1$ for a known signal (the subscript refers to $M = 1$) and for one of $M$ signals, assuming an ideal (optimal) detector. These predicted receiver operating characteristics are shown as solid lines in Figure 2.10 for $d'_1 = 2$ and six values of $M$, ranging from 1 to $\infty$. The optimal detector's stimulus uncertainty loss in detection is a graphic illustration with Normally distributed

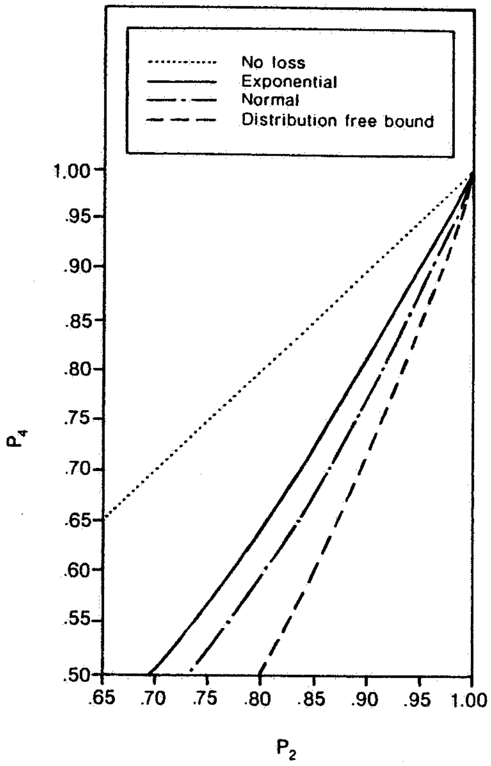



**Figure 2.9.** Theoretical comparison of two versus four alternatives in location judgments. The ordinate shows the predicted probability of a correct location judgment in a four-location task ($P_4$); the abscissa shows the probability of a correct location judgment in a two-location task ($P_2$). The curves illustrate the theoretical accuracy of location judgments of various ideal detectors as the signal-to-noise ratio varies. The curve label indicates the probability density function (*pdf*) assumed for the computation. Values of ($P_2, P_4$) are shown assuming (1) exponential and (2) normal *pdf*s for both targets and distractors, and (3) a distribution-free bound [$P_4 < (P_2)^3$] on of the maximum uncertainty loss in comparing a four-location to a two-location task. The difference ($P_2 - P_4$) is the *uncertainty decrement*. (From M. L. Shaw, Identifying attentional and decision making components in information processing, in R. S. Nickerson (Ed.), *Attention and performance* (Vol. 8), Lawrence Erlbaum, © 1980 by the International Association for the Study of Attention and Performance. Reprinted with permission.)

noise of the general arguments about increasing the number of noise sources given in Section 4.2.1.

Nolte and Jaarsma's (1967) ideal detector, whose performance is illustrated in Figure 2.10, gives equal weight to all channels in forming the likelihood ratio. More generally, when a priori probabilities of signals differ in different channels, or payoffs are unequal across channels, the optimum decision rule requires the likelihood ratio to be based on the appropriately weighted likelihoods in each channel. This is a *weighted decisions* rule, which is considered further in Section 6.2.4. In the optimal decision rule, all the information available from each channel is combined (although the weighting is not necessarily equal) before a decision is made.

Nolte and Jaarsma (1967) were able to show that the optimal likelihood-based receiver operating characteristics could be closely approximated by the behavior of a suboptimum threshold detector that performs a much simpler computation. The threshold detector sets a criterion for each of the $M$ channels separately and responds "signal" when at least one channel or sample exceeds the criterion. When the a priori probabilities and payoffs for the various alternatives are symmetrical, the criterion is taken to be the same in the component channels. Otherwise, different channels have different threshold criteria corresponding to the different weights of the weighted decisions rule. Let $p(\text{“}S\text{”}|E)$ be the probability of a detection response (i.e., criterion exceeded in at least one channel given an input event $E$), and let $p_j(\text{“}N\text{”}|E)$ be the probability of a nondetection response in channel $j$, then

$$p(\text{“}S\text{”}|E) = 1 - \prod_{j=1}^{M} p_j(\text{“}N\text{”}|E) . \tag{10}$$

A threshold detector following the scheme of Eq. (10) is sometimes called a *pandemonium detector* (the loudest channel is heard if its threshold is exceeded), or a *maximum* rule; it differs from the optimal rule slightly. Section 7.2 considers other situations in which analogous rules—separately categorizing each channel as containing signal or noise before combining the channels to make a decision—are optimal.

Receiver operating characteristics generated by the threshold detector as the channel criterion is varied are shown in Figure 2.10 as dashed lines. The predictions of the likelihood-integrating detector and of the maximum detector are practially

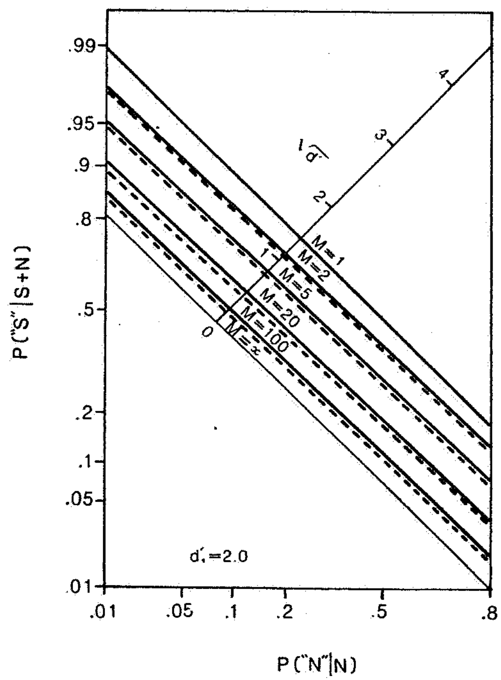



**Figure 2.10.** The uncertainty decrement of an ideal detector as the number $M$ of different, alternative, independent signals increases in a yes-no detection experiment. The computations assume noise has a normal distribution; the $M$ alternative signals are known exactly and each is detected in its own, ideal "channel"; and the signal-to-noise ratio is such that in the most favorable case ($M = 1$, only one possible signal), $d'_1 = 2.0$. The abscissa represents the probability of correctly responding "N" (no detection response in any channel) on a noise trial; the ordinate represents the probability of a correct detection response "S" when a single signal $S_i$ occurs in channel $i$. Solid lines assume an optimal (maximum likelihood) detector that operates on the integrated output of the $M$ channels. Dashed lines indicate performance of a nonoptimal decision maker that applies an optimum detection criterion to each channel separately, and responds "Noise" only when all $M$ channels fail to exceed their individual criterion. (From L. W. Nolte & D. Jaarsma, More on the detection of one of m orthogonal signals, *Journal of the Acoustical Society of America*, 1967, *41*. Reprinted with permission.)

indistinguishable in simple stimulus detection paradigms (as opposed to paradigms in which the stimulus is formed by adding together simple stimuli). However, in multiple-look paradigms (where the subject observes several stimulus presentations before making a response), the likelihood and the maximum detectors yield distinct predictions (M. L. Shaw, 1982). And an adding-of-outputs rule is better than the maximum rule for probability summation experiments (Graham, Kramer, & Yager, 1983; Green & Swets, 1966).

**4.2.4. Attentional Limitations and Uncertainty.** This discussion has focused on the use of compound tasks to investigate attentional *limits* on an observer's ability to perform several tasks (i.e., monitor several locations or frequencies) at once. An attentional limitation would result in poorer information about one or all of the multiple information sources. In the language of SDT, this would be represented by a change in the signal or noise distributions (lower target mean, higher noise variance).

The foregoing discussion reviewed why a loss of accuracy in single-target detection resulting from an increase in the number of locations being monitored is uninterpretable unless one has a theory to determine whether the loss is greater than would be exhibited by an ideal detector. This dependence on a theory (ideal detectors) comes about because the paradigms considered here were exemplars of compound tasks and had more stimulus uncertainty in the compound than in the component tasks, the control conditions. While the complexities of compound tasks may be unavoidable in some real-life situations, they can be avoided in the laboratory by using concurrent tasks which have different, perhaps more tractable, problems. In the case of $n$ locations being monitored, concurrent means that each location has the same probability of containing a target when it is viewed in the context of the other $n - 1$ locations, as it does in isolation. It also means that 0, 1, 2, ..., $n$ targets may occur in a presentation instead of just 0 or 1 as in most compound tasks. Obviously, a large number of targets would pose memory, recognition, and identification problems, as well as detection problems. In multitarget experiments, the occurrence of one target does appear to make the detection of a second target more difficult. These results are described in Section 5.1.1 (see also Gilliom & Sorkin, 1974; Hirsch, Hylton, & Graham, 1982; Pohlmann & Sorkin, 1976; Schneider & Shiffrin, 1977; Sorkin, Pohlmann, & Gilliom, 1973; Sperling & Melchner, 1978b, p. 681). Fortunately, there are paradigms to provide the data to estimate these sources of interference.

Partialing out the effects of stimulus uncertainty from experimental data will affect conclusions about attentional *limitations* on performance that might otherwise have been drawn. In the detection literature, a variety of models (i.e., single band, multiple band, switching single band, etc.) have been discussed extensively (see Swets, 1984, for a review). These "band" models have been called *attention* models in several of the source articles, but they refer to attention in quite a different sense. They do not necessarily reflect *limitations of attention*, but may reflect a *voluntary decision* to weight some sources more heavily and to neglect others. Under appropriate experimental circumstances, electing to sample stimulus information from a narrow-frequency band or from a single location is an optimal strategy to avoid stimulus uncertainty loss, and the neglect of other frequencies or locations does not reflect an attentional limitation. These band models deal with strategic decision processes and not with resource allocation processes. The information from neglected frequency bands or locations may be given less weight (a decision process, e.g., Kinchla, 1977; Kinchla & Collyer, 1974), but the quality of the information (e.g., the signal-to-noise ratio) is not reduced (a resource process).

## 4.3. Visual Search: A Compound Task

**4.3.1. Classical Visual Search Paradigm.** In their classical experiments, Neisser and his collaborators (Neisser, 1964; Neisser, Novick, & Lazar, 1963) studied the ability of the subjects to find a particular target character or characters embedded in long lists of randomly chosen characters. Subjects searched lists from top to bottom and made a manual response when they detected the target. Some sample lists are shown in Figure 2.11. The fastest reported search times were on the order of 20 msec per distractor (nontarget) character. For example, if the target were the thousand and first character on the list, it would take the subject about 20 sec longer to discover the target than if it were the first character on the list. Unfortunately, Neisser and colleagues' calculated search times per character or per row were not consistent between lists having different spatial arrangements of characters.

**4.3.2. Eye Movements in Visual Search.** To investigate the conjecture that eye movements might have been a *limiting factor* in Neisser's visual search, a computer-driven display was devised to enable visual search to proceed without eye movements (Budiansky & Sperling, 1969). In the sequential search procedure, a sequence of briefly flashed letter arrays is presented on a CRT display screen, each new array falling on top of its predecessor. A critical array containing a lone numeral target is embedded somewhere in the middle of the sequence. The target's spatial location (within the array) and its identity are chosen randomly on each trial. The task of the subject is to detect the location and to identify the target (Figure 2.12(a)).

| | | |
|---|---|---|
| HX | VSWO | VSWOGT |
| SP | GTXU | XUHXSP |
| ZP | HXSP | ZPSJXZ |
| SJ | ZPSJ | PZTPGI |
| XZ | XZPZ | VSWOXU |
| PZ | ZPGI | HXSPSJ |
| ZP | QGOB | XZPZGI |
| GI | FIVL | QGOBVL |
| QG | IRDB | IRDBZQ |
| OB | EQZQ | VLSMHT |
| FK | VLSM | HWFBWJ |
| VL | DNHT | DSVEBH |
| IR | HWFB | LFPKHX |
| DB | GWWJ | SUWVYR |
| EQ | DSVE | JBVIWB |
| ZQ | RCBH | BARGCO |
| VL | LFPK | CPLBMR |
| SM | NQHX | PRVNGA |
| DN | SUWV | XYLZBC |
| HT | OTYR | CVYNFM |
| HW | JBVI | WJDVYC |
| FB | SGWB | QGOBFH |
| GW | BARG | VLIRDB |
| WJ | VACO | EQZHVL |
| DS | DPLA | SMDNHT |
| VE | CBMR | HWFBGW |
| LC | PRVN | WJDSVE |

**Figure 2.11.** Sample stimulus arrays of varying configurations for a visual search task like that of Neisser (1963, 1964). The target is "K." Observed search rates (characters per second) vary with display configuration due to eye movements and other factors.

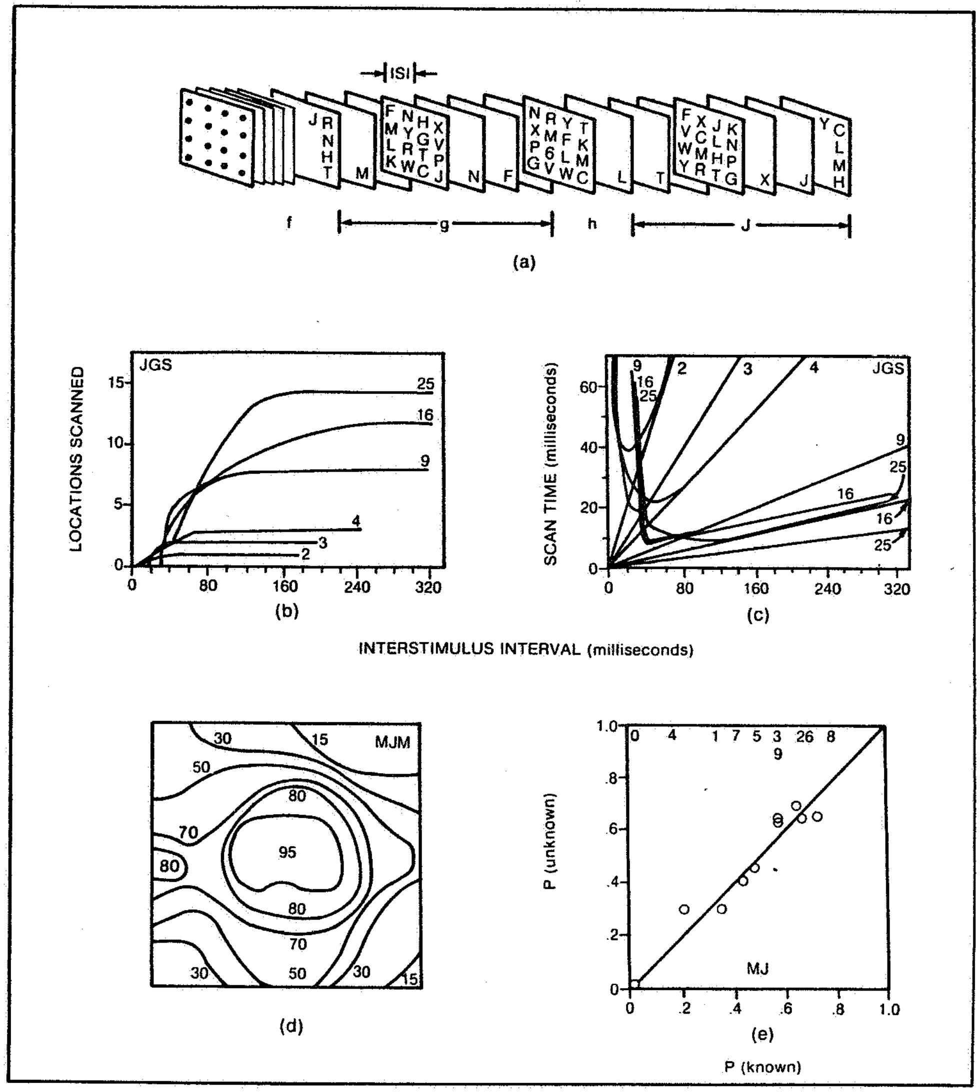



**Figure 2.12.** Computer-generated displays permit rapid visual search without eye movements. (a) The stimulus fixation (*f*); a random number (6, ..., 12) of displays containing only letters (*g*); critical display containing a numeral target (*h*); 12 more nontarget displays (*j*). (b) Number of locations *L* searched as a function of interstimulus interval *ISI*. *L* is corrected for guessing; the parameter is the number of letters in the display. (c) *T*, the scan time per letter, as a function of *ISI*; $T = ISI/L$ is derived from data in (b). (d) Search field contours describing search accuracy at each location of a $7 \times 7$ array. The indicated parameter is search accuracy at the contour. (e) Comparison of the probability correct of a location judgment in search for a known numeral target with search for an unknown 1-of-10 numeral targets. Each numeral at the top of (e) indicates the identity of the target for the data point below. The line through the data has slope 1.0; it accounts for 97% of the variance. Panel (d) is based on Sperling and Melchner (1978b). (From G. Sperling, J. Budiansky, J. G. Spivak, & M. C. Johnson, Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral, *Science*, 1971, *174*. © 1971 by the American Association for the Advancement of Science. Reprinted with permission.)

In rapid, natural visual search through simple material, the eyes make about four saccadic eye movements per sec, each movement lasting a few tens of msec (depending on the distance traversed) with the eyes relatively motionless between saccades (see Woodworth & Schlosberg, 1954, for a historical review). To approximate this natural search mode, the computer-generated arrays are exposed for durations of 200 msec with brief 40-msec blank periods between arrays. The subjects are instructed to maintain stable eye fixation on the center of the display, and they do (Murphy, 1978; Murphy, Kowler, & Steinman, 1975). The successive arrays displayed to the stationary eye approximate the stimulus sequence ordinarily produced by saccadic eye movements. The exposure parameters are not critical. For example, data obtained with 200-msec exposures followed by 40-msec blank periods are not different from data obtained with 10-msec exposures and 230-msec blank periods (Sperling, 1973; Sperling & Melchner, 1976b, 1978a, p. 676).

The computer-generated sequence has many information processing advantages over the natural sequence. For example, in natural search, when the eyes do not move quite far enough between fixations, some of the same material falls within the eyes' search area in successive fixations and is searched twice, which is wasteful. Even if redundant material on the retina is ignored, the redundant material still usurps space within the search area that could have been occupied by new material. If the eyes move too far between fixations, they leave unsearched lacunae in the stimulus.

In natural search, there are two unknown factors: (1) the eye movement strategy and (2) the attentional field around the eye fixations. Eye movement strategy must be known to determine the attentional factors. In the computer-generated sequence, eye movements are effectively eliminated, so that the attentional field around fixation can be determined.

**4.3.3. Maximal Search Rates.** In experimental investigations of visual search in computer-driven visual displays, Sperling, Budiansky, Spivak, and Johnson (1971) studied visual search with many different presentation rates in addition to those that most closely approximated natural search. They discovered that the most rapid visual search actually occurred when new arrays were presented every 40 msec, 5 times faster than the fastest possible saccade rate. See Figures 2.12(b) and 2.12(c). At these artificially high presentation rates, search proceeded at a rate of 1 background character per 10 msec, about twice as fast as Neisser's maximum rate and twice as fast as in the 240-msec presentation rate that simulated Neisser's conditions. In fact, there was only a small difference in detection accuracy between interarray times of 120 and 240 per sec, as shown in Figure 2.12(b), suggesting that in some natural searches the motor control of the eye is indeed the limiting factor. In Neisser and colleagues' search task, if their subjects' eyes had executed saccades every 120 msec, search rates might have doubled with little loss of accuracy. The data of Figure 2.12(b) suggest that the second half of many fixation pauses may have been wasted waiting for the eyes to move.

In contrast to Neisser's lists, the computer-generated arrays of different sizes are searched at similar rates (characters per sec). Further, there is a considerable trade-off possible between scanning characters in one array or in several; thus almost as many background characters can be scanned in one array presented for 120 msec (12) as in 3 arrays, each presented for 40 msec (4 per array). This is best seen by looking at the scan times per character in Figure 2.12(c), which dip just below 10 msec per character throughout the 40–120-msec range of interarray times.

The effective search field around fixation is defined by the proportion of targets detected at various points within it. It is illustrated in Figure 2.12(d) for search of 7 × 7 letter arrays. The search field is approximately concentric, centered slightly above fixation. However, locations with fewer neighbors or with adjacent blank space are easier to search (Bouma, 1978; Harris, Shaw, & Bates, 1979; P. Shaw, 1969) so that the measured search field is distorted by the boundaries of the 7 × 7 stimulus. The subject in Figure 2.12(d) tends to concentrate the search more in the left than in the right half of the stimulus. The search field depends on the stimuli used to measure it; extremely rapid presentations or extremely small size characters shrink the search field. However, these parameter variations do not necessarily alter the *shape of the search field* which suggests that, except for a task-dependent monotonic transformation, the search field is an invariant property of the visual system. Obviously, the search field pattern in part reflects perceptual limitations. Later work has shown, however, that the spatial distribution of attention can be voluntarily altered (Sperling & Melchner, 1978a), so that the search field also reflects voluntary, cognitive factors.

**4.3.4. Application of Detection Models.** Search experiments are variants of the multiple-location detection experiments considered in Section 4.1. Instead of discriminating a low-intensity flash (signal) from background illumination (noise), subjects must discriminate a particular target (e.g., the character 3) from an array of distractors (e.g., C, F, H, R). The underlying decision axis represents "3-ness," where the distractors are drawn from a distribution with a lower mean on this dimension than the target. In Neisser's classical version of the search task, not only is eye movement time the limiting factor in performance but also the functional size of the array and array overlap are unknown. The computer-generated displays solve these problems but cannot solve the theoretical problem of estimating limitations in performance due to various sizes of arrays. Accounting for the effect of array size in visual search is a difficult, incompletely solved problem because at least three factors are usually involved: stimulus uncertainty, retinal nonhomogeneity, and attentional strategy. For theoretical attempts see Fisher (1982), Rumelhart (1970), and Shiffrin and Schneider (1977).

**4.3.5. Visual Search for Multiple Targets: A Compound Task without Stimulus Uncertainty Deficit.** Among the most interesting questions relating to attention in visual search is whether a subject can search as efficiently for one of several targets, say, any of the numerals 0,1, ..., 9 as for a known target 5. This problem has generally been approached by comparing performance in the known-target conditions (where the subject knows the target will be 5) to the unknown condition. According to the analysis earlier in this chapter, the unknown condition is the compound combination of 10 component tasks (where search for the numeral *1* is Task 1, etc.), since only one of the targets appears on any trial. Neisser (1964; 1966) and Neisser, Novick, and Lazar (1963) claimed that subjects could search as quickly for an unknown as for a known target, but they did not test the hypothesis correctly.

**4.3.6. Sequential Search Procedure.** A correct test of the hypothesis that search proceeds as quickly for an unknown as for a known target requires comparing performance for the *same* target in known and in unknown conditions. In the sequential search procedure, comparing detection accuracy for known and unknown targets requires comparing accuracy of the location judgments (*where* in the critical array did the target occur?) in the two conditions. A typical numeral-known condition is a block of 100 trials in which only the target 5 occurs. The corresponding numeral-unknown condition is a mixed list of 1000 trials in which the numeral targets 0,1, ..., 9 occur with equal probability. From this mixed list, the subset of 100 trials, on each of which the target 5 occurred, is extracted for comparison with the known condition.

**4.3.7. Methodological Refinements.** Obviously it would make no sense to compare identification responses between numeral-known and numeral-unknown conditions since the subject knows in advance the identity of the target in the numeral-known condition. However, location judgments can be used to efficiently compare numeral-known and numeral-unknown conditions. A second (but far less efficient) method of comparing detection of known and unknown targets would be to include catch sequences in which no target is present and to determine how well the subject can discriminate catch from target-containing sequences (e.g., compute a $d'$ measure). A third method requires the subject to respond within some brief fixed interval after the critical array to indicate that the target has been detected. Method 3 should be combined with methods 1 and 2, if used. A fourth refinement is to require that the subject give a confidence rating (e.g., "certain," "probable,") of response. This confidence rating can be used to improve the estimate of $d'$ (method 2). Or it can be used to improve the estimate of the probability of a correct location response (method 1). For example, location responses in the lowest-confidence category ("unsure") frequently are found to be statistically independent of the stimulus; that means the subject knows when a response is a guess, and these guesses can be treated differently from other responses.

**4.3.8. Results.** Location judgments were used by Sperling, Budiansky, Spivak, and Johnson (1971) to compare target-known and target-unknown conditions in a visual search task. Accuracy of location judgments for each of the numerals 0,1,...,9 was measured in target-known and target-unknown blocks; these two measured accuracies were nearly the same and were correlated 0.97, as seen in Figure 2.12(e). This near-perfect correlation is very strong evidence that the same search processes were executed in target-known and target-unknown conditions. If there were any substantial differences in search processes, one would expect a somewhat different ordering of target difficulties in the target-known and target-unknown conditions. That is, a target that is relatively easy in one condition might be relatively hard in another. Since this did not happen, it suggests that the search processes in the two conditions are essentially the same process.

Since the target-unknown condition is a compound task, a decrement in search performance would have been attributable to *either* uncertainty loss *or* attentional limitations, and the conclusions would have depended on an uncertainty model. However, since virtually no performance decrement is observed, it can safely be concluded that there is no attentional decrement associated with this particular task combination. Apparently, the 10 searches for each of the 10 alternative targets can be carried out simultaneously, in parallel, without loss.

**4.3.9. Information Overload.** What distinguishes the lossless compound search for an unknown 1-of-10 numerals from the many other similar compound detection tasks that do show stimulus uncertainty losses? First and foremost: *overlearning*. Neisser, Novick, and Lazar (1963) and Schneider and Shiffrin (1977) studied the temporal course of learning as a compound search for arbitrary combinations of targets. Both laboratories found that, with thousands of trials of practice, the initially slow compound search becomes as quick as the search for a single target, which also improves substantially. Following LaBerge (1975), Shiffrin and Schneider labeled the practiced search *automatic* search and observed some interesting properties. For example, when paired as a concurrent task with other tasks, automatic search produces little loading (uses few mental processing resources), and when told to ignore previously overlearned targets, the subject is unable to avoid detecting them. However, these explanations deal with the difficulties of compound search from the human's standpoint. Resource limitations do not restrict an ideal detector, and it nevertheless shows a loss in compound tasks. Why do humans not show more of a loss in this compound-search task?

The second significant aspect of Sperling and colleagues' (1971) sequential-search procedure is that it induces an enormous overload of information. For example, the maximum-likelihood detection model, when confronted with Sperling and colleagues' multiple-array stimulus, would have to compute the likelihood of each of the 10 targets at each of 9 stimulus locations in each of the 10 or more arrays in which the target might occur. About 1000 additions (or more, if target identity and location are to be retained) would have to be computed within a second. Clearly, Nolte and Jaarsma's (1967) procedure of setting a threshold in each channel and responding only if any of the thresholds is exceeded is a more practical alternative. Because there are so many locations and arrays to be searched, thresholds have to be set very high to avoid multiple responses—so high that essentially complete target identification is required. A high threshold does not, without further assumptions, explain why the threshold is so little changed between target-known and target-unknown conditions. But it does suggest that processes different from those proposed for detection of weak signals may be predominant. For example, Posner, Snyder, and Davidson (1980) and others have asserted that the signal detection model is not applicable to all tasks involving above-threshold stimuli. Unalloyed SDT does not explain all detection experiments.

**4.3.10. Location Uncertainty.** When detecting a target, does the subject necessarily know the spatial location where it occurred? A casual examination of the data indicates that the answer is obviously not. Correct identification responses frequently are associated with seemingly random location responses. However, these mislocated detections might have resulted from the subject mistakenly identifying one of the distractors as the target (or from random guessing) and being correct by chance. To separate accidentally correct detection responses from true detections, the refinements of Section 4.3.7 are necessary. Sperling (1984) found that when the identification responses in 5 × 5 arrays were correct and made with one of the top two (of five) confidence ratings, the overwhelming majority of location responses were correct and, more significantly, when occasional errors did occur, 95% of the incorrect location responses were assigned to a vertically or horizontally (but not diagonally) adjacent cell. In this search task, whenever detection occurs, it is associated with a spatial location. Different results have been reported for other search tasks, and when all the methodological precautions have been taken, it would be interesting to know what distinguishes the two kinds of tasks.

## 5. RESOURCE SHARING AND CONCURRENT TASKS

A concurrent combination of two or more subtasks requires the subject to perform all subtasks on each trial. While such a procedure is not feasible in all situations, it has the distinct advantage of not requiring a complex computational model of stimulus uncertainty to place a lower bound on optimal performance. The feasibility of the concurrent task combination often can be determined by examining performance in control conditions. This section reviews the resource sharing model of attentional sharing in concurrent task paradigms, treats several examples of concurrent tasks, and then examines concurrent visual search experiments and the attention operating characteristics derived from them.

### 5.1. Simultaneous Auditory Two-Channel Detection

Consider a typical, compound, auditory detection experiment. In the compound task, any of three possible stimuli occurs within a block of trials: one of two tones differing in frequency or a noise stimulus. The stimuli for component tasks are ($N_1$, $S_1 + N_1$) for Task 1 and ($N_2$, $S_2 + N_2$) for Task 2. Generally, $N_1 =$

$N_2$. A trial of the compound task consists of a random selection of a trial from Task 1 *or* Task 2. A performance loss in the mixed block indicates an inability to share attention between different frequency ranges but only if the alternative hypothesis of a statistical *stimulus uncertainty* loss can be excluded.

**5.1.1. Concurrent Detection.** The concurrent version of the two-tone task presents a stimulus from Task 1 *and* a stimulus from Task 2 on each trial. That is, the appropriate stimulus set for the concurrent task, where the subscript indicates the task, is [$N_1$ & $N_2$], [$N_1$ & ($N_2 + S_2$)], [($N_1 + S_1$) & $N_2$], [($N_1 + S_1$) & ($N_2 + S_2$)], where $S_1$ & $S_2$ indicate joint presentation of $S_1$ and $S_2$. The appropriate response set is ($N_1$ & $N_2$), ($N_1$ & $S_1$), ($S_2$ & $N_2$), ($S_1$ & $S_2$).

In auditory detection, Pastore and Sorkin (1972), Sorkin and Pastore (1971), Sorkin, Pastore, and Pohlmann (1972), and Sorkin, Pohlmann, and Gilliom (1973) use paradigms that are interpretable as the concurrent task just described. Consider, for example, the experiment of Sorkin, Pohlmann, and Gilliom (1973) that investigated the detection of simultaneously presented auditory signals of different frequency. In one of their conditions, stimuli consisting of either 0, 1, or 2 tones, plus noise, were presented to the left ear. Tone 1 was 630 Hz, Tone 2, 1400 Hz. The subjects responded to each tonal stimulus independently; thus, the response alternatives were [0, 1, 2, 1&2]. Sorkin and colleagues (1973) found an interference effect of the concurrent task; detection of Tone 1 (and of Tone 2) was less accurate in the concurrent condition than in the corresponding isolated control task. Additional analyses showed that concurrent detection performance was especially impaired on trials when both tones were present. Given the fact of these performance deficits in a concurrent task, can they be attributed to an attentional failure (the subject is unable to monitor two channels), to response interference (while the subject is responding to the stimulus in one channel, events in the other channel are forgotten), or to some other kind of interference? Further analysis of the procedure reveals two problems taken up in order: (1) the subjects may not be able to *identify* weak stimuli after they have been detected and (2) the procedure is not truly concurrent.

**5.1.2. Discriminability in Concurrent Tasks.** Consider the following compound and concurrent Gedanken experiments. In a detection task, on each trial, the subject is presented with one of two tones or with noise. The tonal frequencies are 630.000 and 630.001 Hz. This is a compound experiment in which performance on the compound task (mixed list of two frequencies and noise) is guaranteed to be equal to performance on the component task (pure list of one frequency and noise) because these two "different" tones will not differentially affect any human performance, if indeed the differences could be physically measured in a brief trial interval. Consider the same three signals in a concurrent paradigm. Either 0, 1, or 2 frequencies are presented on a trial, and the subject must answer separately whether each was present. Designating the signals as ("low," "high") and the absence or presence of a stimulus as 0, 1, respectively, there are four possible stimulus combinations on a trial: (0, 0), (0, 1), (1, 0), (1, 1) and the four corresponding responses. In the concurrent task, the subject must be able to identify the stimulus to perform well. Since the subject cannot do this when stimuli are indiscriminable (0, 1), (1, 0), there is a performance deficit in the concurrent task. The conclusion is that a concurrent procedure makes sense only when the subject knows which task he or she is performing, and concurrency becomes problematical when the component tasks are confusable.

Sorkin and colleagues (1973) do not know how discriminable their tonal signals are, and therefore their concurrent procedure by itself is questionable. Their observation that detection of two binaurally presented tones is not better than detection of the same two tones presented monaurally suggests either that the 630- and 1400-Hz tones are quite discriminable at threshold (or merely that routing the threshold tones to different ears does not make them more discriminable than presenting them to the same ear). But *discriminability* can be determined experimentally, for example, in a two-interval forced-choice procedure in which a 630-Hz tone occurs in one interval and 1400-Hz tone in the other. *Attentional limitations* can be measured directly by attentional manipulations (instructions, payoffs, a priori probabilities) that direct attention to one or the other stimulus. *Cross-stimulus interference* can be estimated by varying the strength of concurrent stimulus events. *Response interference* can be measured by varying the order of report, by comparing partial reports with full reports, by varying the numbers of responses required, and so on. The point of this analysis is that one procedure is not sufficient to resolve a difficult issue, such as multiband auditory detection. That requires both a carefully formulated theoretical framework *and* many convergent paradigms.

**5.1.3. Noise as an Environmental Feature.** A technical point about Sorkin and colleagues' (1973) monaural, concurrent multitone detection task is that the stimuli of component tasks are not ($S_i + N, N$) but ($S_i$, 0). The noise is a feature of the experimental situation, like the chair and the earphones and, technically, not a stimulus. Regarding noise as a stimulus could lead to the selection of $S_i + N$ from task 1 and $N$ from task 2 to produce $S_1 + 2N$ in the concurrent task, which is obviously much more difficult to detect than $S_i + N$ in the component task. On the other hand, in *dichotic* listening, in which each component task is directed at a different ear, the stimuli of component tasks are ($S_L + N_L, N_L$) and ($S_R + N_R, N_R$) where $L,R$ designate left- and right-ear presentations, respectively. This is not an idle quibble because, in dichotic concurrent tasks, the two noise stimuli are uncorrelated (they are chosen independently), whereas in the monaural concurrent tasks, the noise stimulus for each of the component tasks is the same. Technical issues of this sort often are critical to a theoretical understanding of an experiment, and careful analysis of the paradigm in terms of concurrency and compounding may help the experimenter to resolve them. The question of attentional loss in frequency monitoring is still unresolved.

## 5.2. Shadowing

One of the most studied concurrent tasks in attentional research is auditory shadowing (Cherry, 1953; Cherry & Taylor, 1954). The typical shadowing task requires a subject to repeat a message heard in one ear while another message is being presented to the other ear (Treisman, 1964). Early single-channel models of attention (Broadbent, 1958; Craik, 1948) were supported by the observation that shadowing a message in one ear prevents a subject from remembering anything about the content of the other ear's message (Glucksberg & Cowen, 1970; Moray, 1959; Mowbray, 1964; Norman, 1968). One explanation of the subjects' recall failures might be that auditory shadowing of a single message requires most or all of the subjects' attention and that *any* competing task suffers severe disruption. Allport, Antonis, and Reynolds (1972) demolished this simplistic notion about attentional capacity in a series of studies that demonstrated adequate performance on *some* tasks but not others, when performed concurrently with auditory shadowing.

The results of Allport and colleagues' (1972) experiments are shown in Figure 2.13. Figure 2.13(a) illustrates data from three pairs of concurrent tasks. Auditory shadowing was a component task in all task pairs; it was paired with three different recognition tasks, recognition of auditorily presented word lists, visually presented word lists, and visually presented picture lists. The recognition items were selected to yield approximately the same accuracy level when tested alone. In the concurrent task, the subject shadowed the auditory message while attempting to remember the recognition items, which were presented concurrently. Recognition was tested afterward. Quite different levels of accuracy resulted for these three different task combinations. Allport and colleagues (1972) suggest that the degree of compatibility in these task pairs depends on the similarity of the recognition tasks to the shadowing task, which is both linguistic and auditory. Picture recognition, which is neither linguistic nor auditory, is near control levels when performed concurrently with shadowing. Recognition of visual words was impaired by concurrent shadowing, and recognition memory for auditory words (presented in the ear opposite the shadowed message) was impossible.

Allport and colleagues' (1972) experiment exemplifies a common technical error in concurrent experiments: performance on one of the tasks (shadowing) is at ceiling in both the isolated control and the concurrent conditions. Therefore, Allport concludes that sight-reading does not interfere with shadowing. However, it is impossible to know whether there might have been a performance decrement in a more difficult shadowing task, one in which performance was not already at ceiling. Furthermore, an attention operating characteristic must be determined by more than one joint-performance point.

Figure 2.13(b) shows the levels of concurrent performance of auditory shadowing and piano sight-reading. Although both component tasks are difficult, a high level of concurrent performance was achievable simultaneously by Allport and colleagues' (1972) subjects.

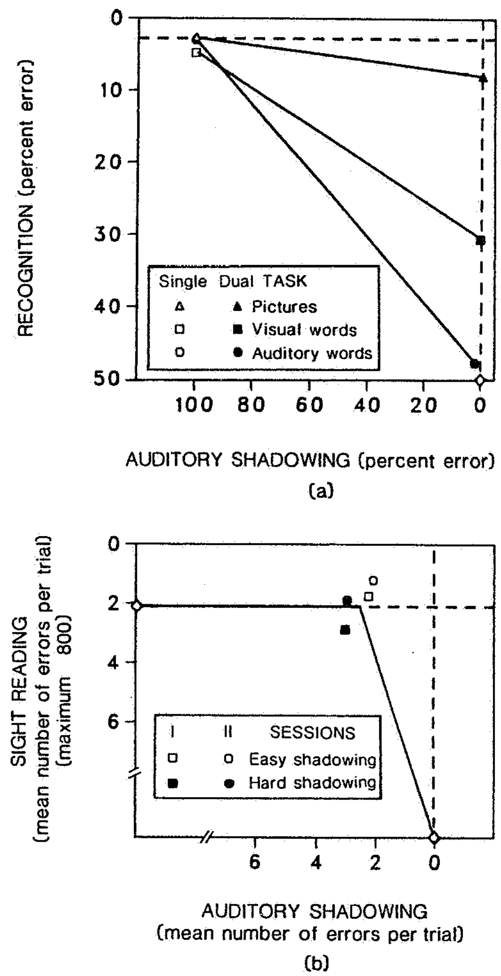



**Figure 2.13.** Results of the concurrent, auditory shadowing experiments of Allport, Antonis, and Reynolds (1972). (a) Accuracy of shadowing performed concurrently with the study phase of a subsequent test of recognition memory for pictures, for visually displayed words, and for orally presented words. Data are estimated from Allport and colleagues' (1972) bar graphs and graphed in a two-dimensional operating space: percentage errors in recognition versus percentage errors in shadowing. The three recognition memory tasks were calibrated for approximately equal single-task baselines (open symbols) but show large differences in the amount of concurrency loss (closed symbols). The mean single-task baseline for the recognition tasks is shown by the dashed line; baseline performance for shadowing is zero errors. This is a problematical design because the shadowing performance is at ceiling. (b) Errors in shadowing and in concurrently performed sight reading of piano music. Diamond symbols indicate the single-task baseline for piano sight reading and for easy shadowing. Concurrent performance data were estimated as in (a); concurrent sight-reading data are separated by levels of shadowing difficulty and practice. Single-task data for difficult shadowing were unavailable. The performance operating characteristic is the light line drawn through the mean single-task and mean dual-task performances.

## 5.3. Concurrent Visual Search

**5.3.1. Arrays Matched to Processing Capacity.** Compound visual search experiments were reviewed in Section 4.3. These experiments (Sperling et al., 1971) were directed at finding the optimal stimuli for visual search, but the stimuli were restricted to one size of character. What is the optimum size of character for visual search? Many small characters can be presented in the foveal area where acuity is good, but small characters are below the acuity limit of peripheral vision. Conversely, composing an array of large characters that are resolvable in peripheral vision causes central acuity to be squandered; the fovea will be fully occupied by a mere fragment of a character. The obvious solution is to compose an array of characters of different sizes. How should characters of different sizes be arranged to facilitate the scan of the largest possible number of characters? Place small characters in the center ranging to large characters in the periphery, where each size of character is matched to the information processing capacity of the retinal area on which it was imaged. Anstis (1974) developed such displays, which he used for demonstrating letters that are equally above their acuity threshold in different areas of the retina.

**5.3.2. Concurrent Search for Large and Small Targets.** The investigation of visual search in arrays that are spatially matched to visual information processing capacity was undertaken by Sperling and Melchner (Sperling, 1975; Sperling & Melchner, 1978b). Array sequences were constructed in which only one target numeral occurred in a critical array otherwise composed entirely of letters. Figure 2.12(a) illustrates the procedure. This target might occur at peripheral locations that received large-size targets or central locations that received smaller targets. Figure 2.14 shows one of several array configurations tested. Surprisingly, Melchner was unable to search arrays simultaneously for large and small targets. That Melchner could not search simultaneously for a large and a small target (e.g., a large *9* or a small *9*) in the same array was especially astonishing since, in earlier experiments, he had been able to search simultaneously, without loss, for 10 numeral targets (0,1,...,9) when they were all the same size. Is a large *9* more different from a small *9* than from a large *3* or *4*?

The appropriate search task to test this possibility is a concurrent search for a large and small target numeral. Figure 2.15 illustrates a display consisting of 16 large characters on the outside perimeter and 4 small characters in the center. The character sizes were adjusted to make detection of the small foveal target approximately as difficult as detection of the large peripheral target. The attentional question concerned the subject's ability to search for both a small and a large numeral *concurrently*. Hence the large and small targets both appeared on the same frame in the search task.

*5.3.2.1.* ***Procedure.*** Sperling and Melchner (1978a, 1978b) presented a long sequence of briefly flashed arrays at a rate of 4 per sec. A critical array embedded in the middle of the sequence contained a randomly chosen numeral target at 1 of the 16 outside locations and another randomly chosen numeral at 1 of the 4 inside locations. In the main experimental conditions, the subjects' task was to detect both targets. The subjects had to state the identity, location, and their confidence level for each of the two targets. The subjects' allocation of attention was explicitly controlled by instruction. In some blocks of trials, they were told to give 90% of their attention to the inside target and 10% to the outside target; in the other blocks the instructions were reversed; and in still other blocks, subjects were instructed to give equal attention to both classes of targets.

*5.3.2.2.* ***Results.*** Some useful methodological innovations were incorporated in the analysis of these data. Responses on which the lowest confidence was used (zero, "guessing") were found by chi-square tests to indeed be statistically independent of the stimuli. This means the subjects know when they do not know (cf. Sperling & Melchner, 1976a, p. 209). Further, analysis of verifiable location errors showed that more than 95% of the time when a target was mislocated, it was mislocated at an adjacent horizontal or vertical (not diagonal) position. These two additional items of information can be incorporated into a rigorous, tripartite criterion for true identifications, namely, (1) correct identification response, (2) confidence greater than zero, and (3) mislocation not greater than one adjacent position.

The data for concurrent search of large and small characters with various attention instructions are shown in Figure 2.15(a). The abscissa represents the percentage of correct target identifications of the outside targets; the ordinate represents the percentage of correct identifications of inside targets. Each data point represents the average of data collected in several blocks of trials. The data fall along a line of slope approximately −1, indicating that as probability of identifying one class of target increases (according to the instructional demand), it is compensated by an almost exactly equivalent decrease in identification probability for the other class of targets. The locus of all achievable joint performances on the two tasks (approximated by the straight-line segments connecting the data points) is the attention operating characteristic (Sperling & Melchner, 1976b, 1978a). The term was proposed by Kinchla (see Kinchla, 1980, p. 217, and Sperling, 1984, p. 112) following the terminology of signal detection theory (SDT) (Swets, 1964).

If subjects could search for both targets concurrently without loss, then their performance in all experimental conditions would fall on the *independence point*—the point at which subjects identify both large and small targets concurrently as accurately as they do in the corresponding control condition. (This is the upper right point of the square in Figure 2.15(a).) Clearly this point was not achieved; there always was some loss.

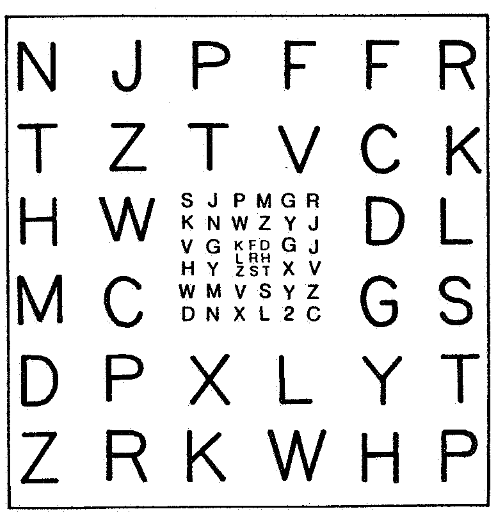



**Figure 2.14.** A search array in which character size has been approximately matched to the information processing capacity of the visual system. The target is a single numeral. This display does *not* maximize performance because subjects have difficulty searching concurrently among characters of different sizes. (From G. Sperling, A unified theory of Attention and signal detection, in R. Parasuraman & R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

### 5.3.3. Performance in the Component Tasks and Other Control Conditions

*5.3.3.1.* ***Control for Memory Overload.*** A series of trials was run in which the letter distractors were replaced by dots. This made target identification trivially easy, and subjects never failed to report both targets correctly. Thus any errors subjects may make in experimental conditions are due to their inability to detect the targets among distractors, not to their inability to report both targets, once detected.

*5.3.3.2.* ***Component Tasks.*** In some blocks of trials, subjects were instructed to report *only* outside targets and to ignore inside targets, and in other blocks, they received the reverse instruction. These control data are graphed directly on the coordinate axes of Figure 2.15(a). That the probability of report is nearly equal in the two control tasks (inside, outside) is not a coincidence; the character sizes and array sizes were chosen to match the tasks in difficulty.

### 5.3.4. Three Concurrent Pairs of Search Tasks.

To gauge the amount of loss in concurrent search, it is informative to investigate several related pairs of concurrent tasks. In all, three pairs of tasks were studied. One task in each pair remained precisely the same throughout: detection of a numeral among the outside letters. Three different inside tasks were matched to this task in difficulty: (1) detection-identification of a small inside target; (2) detection-identification of a normal-size inside numeral (where every inside character was obscured by a ran-

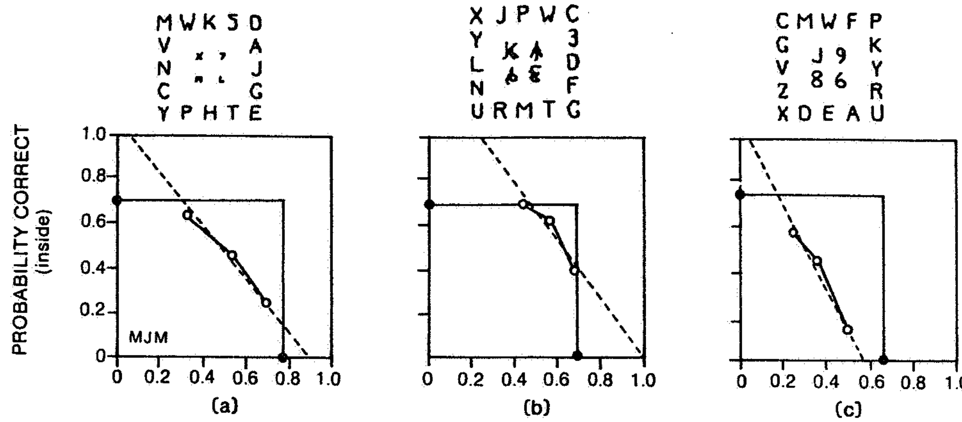



**Figure 2.15.** Stimulus configurations and the empirical attention operating characteristics (AOCs) for three pairs of concurrent tasks. The trial structure in the experiment was similar to that shown in Figure 2.12(a). (a) A large (outside) and small (inside) numeral target appear concurrently for independent detection. Abscissa and ordinate indicate the percentage of correct identifications of the outside target and inside target, respectively. Isolated control conditions are shown as filled circles on the coordinate axes. The intersection of the perpendiculars drawn through these control points defines the independence point. Concurrent performance is indicated by open circles. Attention conditions, ordered from upper left to lower right, respectively, are 90% to inside, equal, 90% to outside, with each point representing the average of two or three blocks of trials. The heavy line connecting the data points is the empirical AOC. The broken line represents the best-fitting straight line to the data. (b) Same conditions as (a) except the inside task is detection of a noise-obscured numeral target of same size as the outside target. (c) Same conditions as (a) except the concurrent inside task is detection of a letter target among three numeral distractors. These three task combinations show different levels of compatibility, as indicated by the distance of the AOC from the independent point. (From G. Sperling & M. J. Melcher, The attention operating characteristic: Some examples from visual search, *Science*, 1978, *202*. © 1978 by the American Association for the Advancement of Science. Reprinted with permission.)

domly chosen "noise squiggle"); and (3) detection-identification of a single target *letter* among inside numerals.

The same control and experimental conditions as before were conducted with these stimuli. Figure 2.15 shows the attention operating characteristics (AOCs) generated by these three pairs of tasks. The distance of the AOC from the independence point is a measure of the incompatibility of two tasks. As in the classroom example of Figure 2.1(c), perfectly compatible tasks, performed as well concurrently as in isolation, would fall on the independence point.

As an index of compatibility between two tasks, we can take the percentage of *isolated* performance achieved by the *concurrent* performance. Concurrent performance is averaged over the component tasks under conditions of equal attention, that is, at the point where the AOC curve—or surface in higher-dimensional space—crosses the line connecting the origin and the independence point. (Let $B$ represent the percent correct measure; $A$, the area under the AOC, is a better but more complex measure than $B$, given approximately—in this example—by $A \approx 1 - 2(1 - B)^2$.) The most incompatible pair of tasks consists of (1) searching for a numeral among letters concurrently with (2) searching for a letter among numerals. These tasks are almost mutually exclusive, average concurrent performance being about 54% of isolated performance. (By doing only one task or the other, never both—even under concurrent instructions—50% of isolated performance would be achieved, by definition.)

The most compatible tasks are (1) searching for a numeral (on the outside) and (2) searching for a numeral of the same size obscured by noise (on the inside). Concurrent performance is about 82% of isolated performance. The original pair of tasks (search for a large and for a small numeral) falls in between with a concurrent performance of 66% of isolated performance. Apparently, searches for a large 9 and a large 3 are more compatible than searches for a large 9 and a small 9. As usual, however, matters are not quite as simple as they first appear to be. When both large and small characters could occur in any of a small number of central locations (rather than being confined to spatially separated areas), Sperling and Harris (Note 1) found no effect of attentional instructions; performance was at the independence point. A similar result was reported by Hoffman and Nelson (1981).

The visual search experiment is perhaps the most complete example of the use of concurrent tasks to investigate attention. The performance levels of each component task (search inside, search outside) were explicitly measured. The concurrent combination was tested under several attention instructions (verbally defined utilities). Appropriate controls were performed to determine that the performance losses were "attentional" and could not be attributed to reporting bottlenecks.

## 5.4. Attention Operating Characteristics

**5.4.1. Determining Entire AOCs versus Determining Single Points.** In the analysis of concurrent tasks previously described, there are two processes by which performance may differ. Different task combinations move performance from one AOC to another, and varying attention allocation between two tasks moves performance along a single AOC. To compare the compatibility of two pairs of tasks, it is necessary to obtain the two AOCs, not just single points on the AOCs. The situation is analogous to SDT in which, to compare the detectability of two signals, two receiver operating characteristics (ROCs) are needed, not just one point on each. In two-task concurrent paradigms where only one point per AOC is measured, for example, one point for task A with task B and another for task A with task C (as in the examples of Section 5.2), one cannot be sure that subjects have applied the same implicit attentional allocation in the two concurrent experiments. Generally, one cannot draw quantitative conclusions, and in some cases even qualitative conclusions may be in error.

**5.4.2. Secondary-Task Procedure.** To illustrate the pitfalls of determining less than a complete AOC, consider the *secondary-task* procedure (Kahneman, 1973), a concurrent paradigm sometimes used to measure attentional requirements of a task. In this procedure, a secondary task *C* is performed concurrently with a primary task *A*. The subject is instructed to optimize performance on the primary task, that is, to hold performance on *A* as close to control levels as possible. It is assumed that the observed level of performance on *C* will provide a measure of the attentional requirements of the primary task *A*. To compare the resource requirements of task *A* to those of task *B*, each of *A* and *B* is paired with *C*, and the corresponding deficits in performance of *C* provide the index for comparison.

The classroom analogy is helpful in conceptualizing the problem in the secondary-task procedure. Examples are shown in Figure 2.16. Subjects are given instructions to operate near 100% of control levels on the primary tasks *A* and *B*. Figure 2.16(a) illustrates a case in which it would be erroneously concluded that these tasks were exactly equivalent in attentional demand characteristics, although they differ substantially. Conversely, in the situation illustrated in Figure 2.16(b), the secondary-task method produces an overestimate of the difference between tasks *A* and *B*. Perhaps these unfavorable hypothetical situations are unlikely. However, as one can safely restrict the number of attentional conditions to be studied only after one knows the AOC and one can know the AOC only by studying several attentional conditions, there seems to be no alternative to measuring AOCs.

## 6. REACTION TIMES AND SPEED–ACCURACY TRADE-OFF

Many investigations of attention employ reaction-time (RT) paradigms. The same task classifications apply to RT tasks as

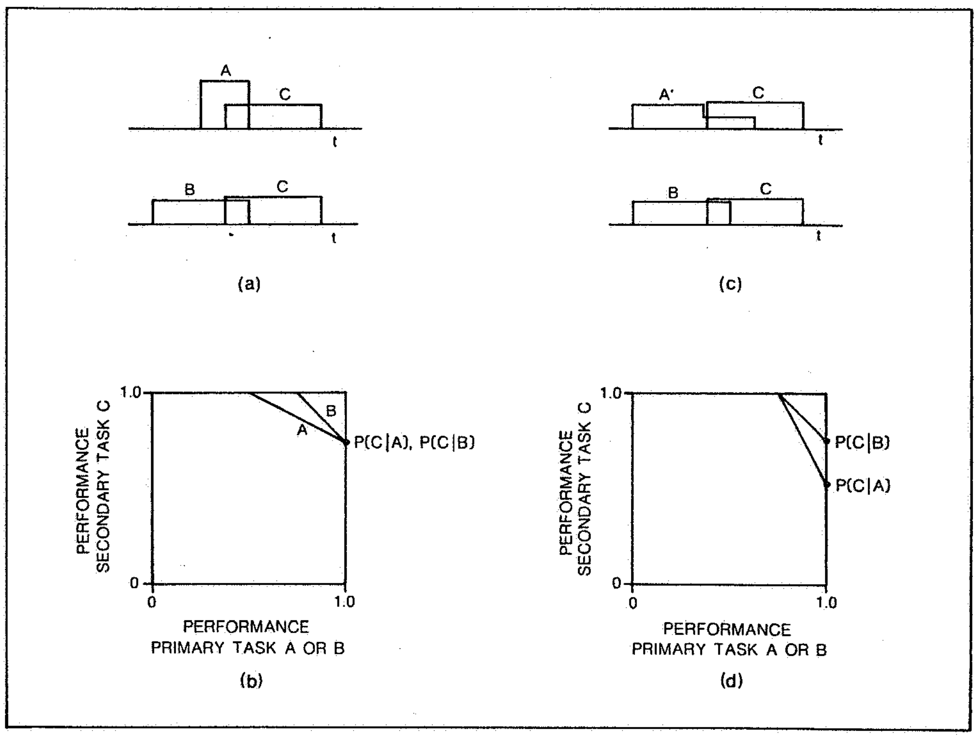



**Figure 2.16.** Representation of the "secondary task" procedure in classroom examples. (a) Information densities for two primary tasks *A* or *B* (classes) combined with a secondary task (class *C*) to yield secondary task deficits $\Delta(C|A) = 1 - P(C|A)$, and $\Delta(C|B) = 1 - P(C|B)$. (b) The attendance operating characteristics (AOCs) corresponding to the task combinations $(A, C)$ and $(B, C)$ in panel (a). The abscissa is the proportion of single-task performance on the concurrent primary task; the ordinate is the proportion of single-task performance on the concurrent secondary task. These different task combinations yield different AOCs but identical secondary-task deficits [shown by the point $P(C|A) = P(C|B)$] when primary task performance is held near control levels, as per instruction in this paradigm. (c) Alternative information densities for two primary tasks $(A', B)$ and a secondary task (*C*) to yield a larger secondary task deficit. (d) Operating space as in panel (b); the AOCs correspond to the information densities in panel (c). With these task combinations, a secondary task experiment yields the secondary task deficits shown by the points $P(C|A)$ and $P(C|B)$. Panels (a) and (b) show that different fractions of overlapping resources (of *A* and of *B* with *C*) can lead to identical secondary-task deficits; panels (c) and (d) show that the same fraction of overlapping resources (of *A* and of *B* with *C*) can lead to different secondary-task deficits, illustrating the danger of using a single point from a secondary-task procedure rather than obtaining a full AOC.

to tasks that use only accuracy measures. This section develops a theoretical approach to ideal (lossless) RT performance tradeoff, simple RT experiments, choice RT experiments, speed–accuracy trade-offs (SATs), and utility functions for some of these procedures.

## 6.1. Ideal Performance in Compound Reaction-Time Tasks: Random Walk Model

**6.1.1. Definition of Reaction-Time Tasks.** In a simple RT task, the subject is presented with a single stimulus and asked to make the designated response as quickly as possible. A choice RT task is the union of $k$, $k \geq 2$ simple RT tasks. Any of the component stimuli $[s_j, j = 1, \ldots k]$ is presented and the subject makes a response, $r_i$. Only $s_j - r_j$ pairs are "correct" (e.g., utility = 1); others are "wrong" (e.g., utility = 0). A choice RT task requires all stimuli and all responses be different. When all the responses $r_j$ are the same, the compound task is sometimes called a *Donders Type 3* task.

**6.1.2. Examples.** Consider the following kinds of RT experiments. On each trial, a subject is presented a stimulus which must be classified into one of two (or more) categories as quickly as possible. For example, the subject may be shown a letter string and be asked to press a reaction key with the left hand if it is a word or another key with the right hand if it is not. This is the lexical decision task originally described by Rubenstein, Garfield, and Millikan (1970) and Rubenstein, Lewis, and Rubenstein (1971a, 1971b). Or the subject may be asked to classify a colored patch that has an irrelevant color name written on it, as red or green. When the color of the patch and the color name differ, the subject is slower and less accurate than when the name is omitted. This is called the *Stroop effect* (Kahneman & Treisman, 1984; Stroop, 1935). Or the subject may be asked to classify stimuli by means of a card-sorting task, placing cards as quickly as possible into different piles according to category. All these tasks are compound (not concurrent) tasks; the subject is presented only one of the possible stimuli and makes only one of the alternative responses on any one trial. RT tasks usually do not involve multiple or concurrent responses to concurrent or simultaneous stimuli, although there are exceptions, for example, the double-stimulation paradigm (Karlin & Kestenbaum, 1968; Smith, 1968). For a review of the double-stimulation paradigms, see Kantowitz (1974). To determine whether an attentional loss occurs in a compound RT experiment (as compared to a simple RT control experiment), it is useful to compare performance in compound RT experiments to a model of an ideal (or lossless) processor.

**6.1.3. Random Walk Model.** The *random walk model* of Link (1975) and Link and Heath (1975) is a simple theory for RT closely related to signal detection theory (SDT). SDT is a model for the perception and decision component in detection tasks. The random walk model is a theory for the perception and response–decision component in RT tasks. (For a review of other possible models, see Green & Luce, 1973.) Without going into full detail, the principle of the random walk model can be summarized as follows: an ideal (lossless) detector accumulates information from the start of a trial, and when the information exceeds a threshold, the appropriate response is made. Each new increment of information is assumed to be somewhat unreliable, so that the cumulative balance of all the information may waiver between the alternatives, that is, execute a random walk. A strategy consists of a choice of *response threshold* (the distance from the starting point to the absorbing boundary) for *each* of the alternative responses. Equivalently, Link's (1975) notation uses the distance between the two boundaries, $A$, and a bias, or starting point between them, $C$.

Random walk models are a subclass of the larger class of *sequential decision strategies*, characterized by (1) continuing to make observations until, (2) some function on the sample space of observations is satisfied, and (3) making the decision. There is a wide range of problems for which functions are known that will make a sequential strategy the optimum strategy (Wald, 1950). The random walk model, as outlined, is known to be an ideal detector—that is, an optimum strategy—when the incoming information can be regarded as symmetric between the choice alternatives (Laming, 1968). For example, evidence is like the outcome of a toss of a coin biased 0.55 in favor of one side or the other; a decision in favor of a heads-bias would be made when the number of observed head tosses exceeded the number of tails by a criterion and vice versa for a tails-bias decision.

To optimize its performance with respect to the experimentally defined payoffs, the response threshold of the random walk model is adjusted so that an optimum compromise is made between several incompatible criteria. The response threshold is set high to avoid accidental incorrect responses (due to some randomness in the incoming information) but not so high that the RT is too long. (The higher the threshold, the longer it takes, on the average, to reach it.) These relations are illustrated in Figure 2.17. A priori information that a stimulus is probable will cause the threshold for the corresponding response to be set lower, thereby decreasing RT and increasing the accuracy when that stimulus is presented, and decreasing the accuracy when the other stimulus is presented. A priori information that a stimulus is unlikely forces the response threshold to be raised in order to avoid errors. The response threshold is changed by changing $A$ or $C$ or both together. Several sample random walks are shown in Figure 2.17.

In SDT attentional *limitations* were represented by change in the underlying distributions (lower signal mean or higher noise variance) as a result of an attentional manipulation. In an ideal detector, decrements in performance occur with increasing numbers of potential stimuli; thus corresponding decrements in human performance might still be compatible with a lossless detector. Analogously, in the random walk model, attentional limitations are represented by changes in the *quality* of the accumulating information—the rate of growth of internal $d'$ with time. Changes in expectancies may alter the criteria (response thresholds, biases) in the random walk model, but they do not reflect a limitation in attentional capacity. When the experimental conditions are varied, in the random walk model of choice RT, as in SDT, the issue becomes one of attributing the corresponding performance variations to criterion changes or to sensitivity changes.

## 6.2. Costs and Benefits in Reaction-Time Tasks

The cost-benefit paradigm is an example of an expectancy manipulation that has been interpreted within an attentional framework. For reviews of experiments, see Audley (1973) and Welford (1980a, 1980b). Consider the following experiment by Posner, Nissen, and Ogden (1978). A subject views a fixation point between two locations, designated "left" and "right," where a light flash may appear on a given trial. Whichever flash ap-

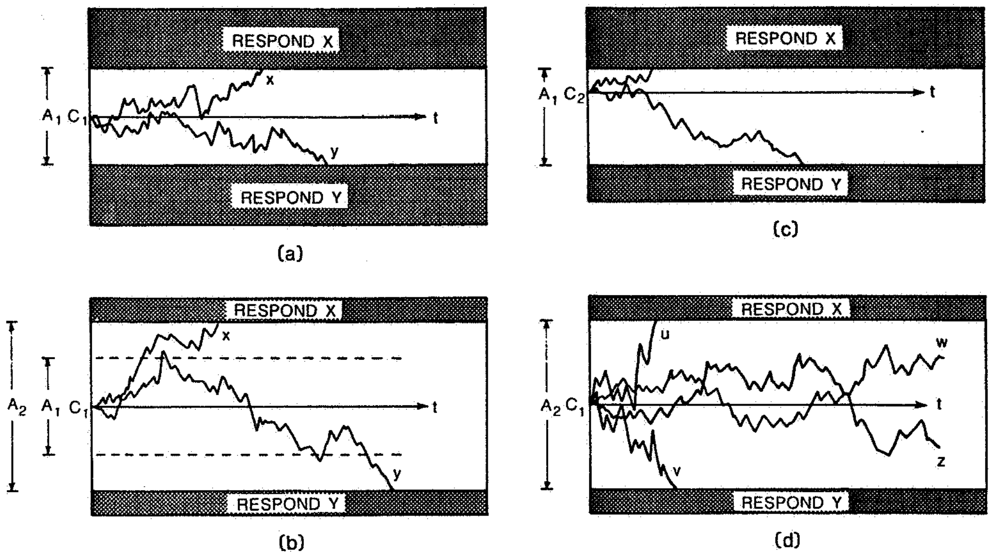



**Figure 2.17.** Graphic examples of the role of the parameters in Link's (1975) random walk model (RWM). (a) Illustration of a RWM for a choice RT with two alternative stimuli and corresponding responses. The abscissa represents time *t*; the ordinate represents the position (cumulative information value) of the random walk at time *t*. The random walk executed on any given trial is statistically dependent on the stimulus. Sample random walks leading to correct responses *X* and *Y* are labeled with the stimulus value (*x* or *y*). When a random walk reaches the response boundary a RT response is initiated. Expected accuracy and speed of the model are controlled by the sensitivity parameter *A* (the sum of the threshold distances) and the bias parameter *C* (the starting point). The starting point is adjusted to reflect an expectation of stimulus *x* versus stimulus *y*. (b) The same random walk processes as (a) but with greater distance $A_2$ between response boundaries. This leads to longer expected RTs and higher expected accuracy than for panel (a). The random walk to stimulus *y* would have resulted in an error with response boundaries $A_1$, but not with response boundaries $A_2$. (c) Same random walk processes as in panel (a) but with high a priori expectation of stimulus *x*, reflected in a lower threshold for response *X* than response *Y* (shift of starting point $C_2$ toward boundary *X*). Starting point $C_2$ toward boundary *X*). Starting point $C_2$ leads to very fast correct responses to stimulus *X* but also to frequent fast errors when *Y* is presented. (d) The effect of speeding up $(u, v)$ or slowing down $(w, z)$ the drift rate of random walks (relative to $x, y$ in panel (b)) is different than changing threshold parameters *A* or *C*. Speed reflects the *rate* at which information is acquired; *A* and *C* reflect the *accuracy* and *bias* parameters of the decision rule that determines the response, given the available information.

pears, the subject is to respond as quickly as possible by pressing a key. Occasional blank trials (no flash) are introduced to reduce anticipatory responses (responses before the flash). This is a *go/no-go* RT experiment in which the subject must respond ("go") when any stimulus is presented and must not respond ("no-go") on catch trials. The experimental manipulation of concern here is the fraction $\alpha$ of stimulus-containing trials on which the left stimulus appears. Posner and colleagues (1978) investigated three conditions: trials in which $\alpha$ was, respectively, 0.80, 0.50, and 0.20. Trials with different $\alpha$'s traditionally are run in separate blocks. However, in Posner and colleagues' experiment, these conditions are interleaved in a mixed-list design; a warning cue (1 sec before stimulus presentation) informs the subject of $\alpha$.

Posner and colleagues' experiment is a two-task compound experiment in which the two component tasks are (1) press the key when the "left" flash appears and (2) press the (same) key when the "right" flash appears. One dependent measure in Posner and colleagues' experiment, as in virtually all RT experiments, is *mean* RT. Ignore, for the moment, the other dependent measure, *accuracy*, which, in this experiment, is determined by errors that occur when the observer responds before the stimulus occurs (or within 100 msec of its onset), fails to respond within a reasonable time period, or responds on a catch trial. The observed RTs for each of the two component tasks in each of the three conditions is represented in Figure 2.18(a). [Except for a slowing of RT, Posner et al. (1978) found no important differences between the RTs in this Donders Type 3 RT experiment and in a choice RT experiment in which the subject had to press a left key in response to the left flash and a right key to the right flash.]

The data from Posner and colleagues' experiments also can be graphed in operating space; thus the data of Figure 2.18(a) are graphed as an operating characteristic in Figures 2.18(b), 2.18(c), and 2.18(d). Fast reaction time represents good performance, and in this chapter we maintain the convention of representing good performance up and to the right.

**6.2.1. Iso-utility Contours for Reaction Times.** What are the utilities in Posner and colleagues' experiment? The authors did not define these explicitly for the subjects. However, suppose that utility varies in direct inverse proportion to the RT: the faster the reaction time, the higher the utility. With this assumption, the expected utility *EU* of any joint performance [*RT*(*left*), *RT*(*right*)] can be computed as a function of $\alpha$, the proportion of left stimuli and the mean RTs:

$$EU = -[\alpha RT(left) + (1 - \alpha) RT(right)] . \quad (11)$$

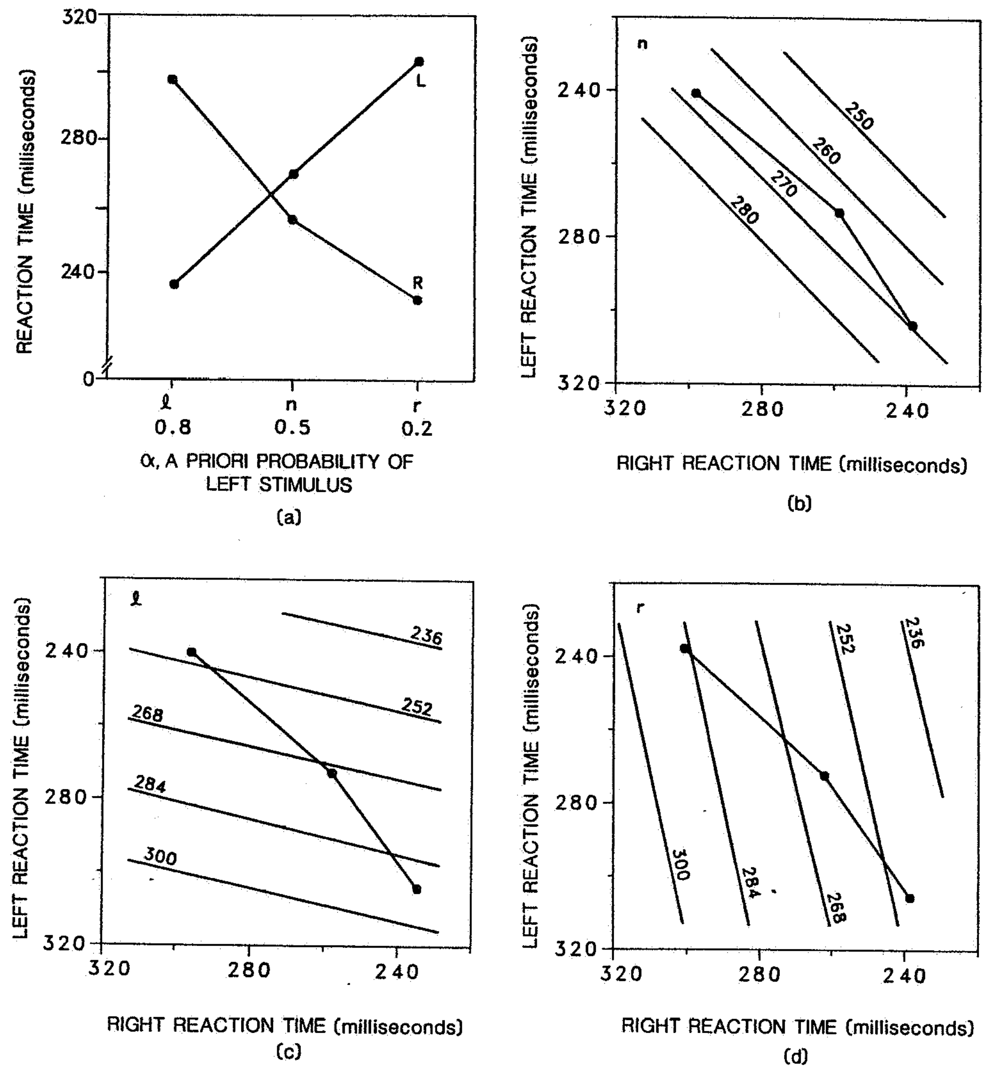



**Figure 2.18.** Iso-utility contours and trade-offs in RTs to stimuli with different a priori stimulus probabilities. All panels illustrate the same data from an experiment by Posner, Nissen, and Ogden (1978). One of two alternative stimuli (a left *L* or right *R* stimulus light) was presented on each trial; the response (a key press) was the same in either case. Three different probabilities α of the left stimulus were tested; the subject was informed of these probabilities by three warning cues (left *l* for $\alpha = 0.8$, neutral *n* for $\alpha = 0.5$, and right *r* for $\alpha = 0.2$). (a) Conventional representation of reaction times to the two stimuli (*L*, *R*) with the three a priori probabilities α. The ordinate is mean RT in msec. The abscissa is α, the a priori probability of the left stimulus as indicated by the warning cue. In panels (b), (c), and (d), the data of (a) are regraphed as a performance operating characteristic; the panels differ only in the iso-utility contours. The coordinates represent RTs to the left and right stimuli, respectively, and are oriented to show good performance up and to the right. The iso-utility contours in panels (b), (c), and (d) represent the mean RTs indicated on the contour; that is, each point along the contour represents a joint RT performance to left and right stimuli; the overall mean reaction time is the *u* value indicated on the contour, $u = \alpha RT_{\text{left}} + (1 - \alpha)RT_{\text{right}}$. The iso-utility contours represent a weighting of performance appropriate to the stimulus probabilities for the conditions. (b) Neutral cue *n*, $\alpha = 0.5$; (c) Left cue *l*, $\alpha = 0.8$; (d) Right cue *r*, $\alpha = 0.2$. (From G. Sperling, A unified theory of attention and signal detection, in R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

Expected utility happens to be a negative number; it increases (approaches zero) as performance improves (RT becomes faster). Iso-utility functions based on Eq. (11) are illustrated in Figures 2.18(b), 2.18(c), and 2.18(d) for the three values of α for which data are available. Note that the utility functions derived from Eq. (11) are similar to those in typical signal detection tasks, but the data are not, in that the data approximate a straight, rather than a curved, line. Straight-line data in this experiment, as in Sperling and Melchner's (1976b, 1978a, 1978b) attention study, have special significance; they suggest that a mixture of just two states, rather than a continuum of states, is sufficient to account for the data. This point will be taken up in detail later in Sections 8.2, 8.6, and 8.7.

Posner and colleagues' (1978) observers seem to operate sensibly with respect to the utility function Eq. (11), optimizing their performance in each case. See Figures 2.18(b), 2.18(c), and 2.18(d). Knowing $\alpha$ when $\alpha \neq 0.5$ enables the observers to shorten their mean RT substantially over the mean RT when $\alpha = 0.5$. The "benefit" to RT (i.e., the reduction of RT in msec) when the more probable stimulus occurs is about the same as the "cost" (RT increase) when the less probable stimulus occurs. The important point is not that the costs and benefits of an asymmetric a priori stimulus probability $\alpha = 0.8$ happen to be approximately symmetric but that the benefits are available on 80% of the trials, while the costs are incurred on only 20%. Thus the mean RT improves with asymmetric stimulus probabilities in a way completely analogous to the improvement of *S/N* detection accuracy with asymmetric stimulus probabilities, considered in Section 4.2.

A second important point about RT benefits is that, once it is known that a simple reaction time $RT_S(i)$ to a stimulus $i$ is quicker than the reaction time $RT_{CP}(i)$ to the same stimulus embedded in a compound RT task, then RT benefits follow directly as a consequence of the procedure for measuring them. To see this, let $\alpha$ represent the fraction of trials of stimulus $i$ in the compound mixture. Then

$$\lim_{\alpha \to 1} RT_{CP}(i,\alpha) = RT_S(i) \ . \tag{12}$$

Equation (12) follows because, as $\alpha \to 1$, the *physical* descriptions of the tasks represented by the left and right sides of the equation become identical. In the limit then, as $\alpha \to 1$, the RT benefit approaches the limit set by the previously determined simple RT. The magnitudes of the RT costs for the tasks that occur with probability $1 - \alpha$ remain to be determined empirically.

### 6.2.2. Random Walk Models of Reaction-Time Paradigms

*6.2.2.1. Choice Reaction Times.* Choice RT is a compound task. In a choice RT task, a left and a right stimulus would each require a corresponding left- or right-hand response, instead of the same response as in the multistimulus go/no-go task. Random walk models for choice RTs were treated in Section 6.1.3 and illustrated in Figure 2.17.

*6.2.2.2. Concurrent Reaction Times.* A concurrent RT task might require a left-hand response to a left stimulus and a right-hand response to a right stimulus, where both stimuli would appear on some trials. In some respects, this would be the ideal test of whether subjects could deal with two stimuli as well as one, but it might involve additional difficulties if conflicts arose in the motor system (Kantowitz, 1974). Assigning the locus of performance loss to attentional processes (versus perceptual or motor processes) can be complicated, but it is solvable (e.g., Sperling & Melchner, 1978b).

*6.2.2.3. What Constitutes Evidence for Loss in Reaction-Time Trade-offs.* The original question in Section 6.2 was whether the subject was able to deal with two possible stimuli as effectively as with one. For a multistimulus go/no-go task to inform us as to whether the subject is unable to deal simultaneously with two possible stimuli (and hence must divide attention between stimuli according to the a priori probabilities), it is necessary to compare performance in the compound task to a model of ideal (lossless) performance. The compound task in which either a left or right signal, or no signal, occurs on each trial is analogous to the compound signal detection experiment where $S_1 + N$, $S_2 + N$, or $N$ appear on each trial. In SDT an ideal observer should show some loss in combining information about the two signals. Furthermore, an ideal detector should respond to changes in a priori signal probabilities or changes in the payoff matrix by changing the criteria but not by changes in the quality of information upon which a judgment is based. Is it possible, in Posner and colleagues' multistimulus go/no-go experiment, that performance varies with instructions and payoffs and yet the quality of perceived information remains invariant? That is, does an observer react more slowly when there are two locations to monitor because information cannot be processed as efficiently from two as from one location, or does the observer's slower reaction reflect the same loss that an ideal detector with no information loss would show in the same situation? As with all compound tasks, a theory is necessary to answer this question.

The random walk model (RWM) is a model of an ideal detector for reaction time; it is necessary to apply an RWM to Posner's task to answer the original question. However, to apply an RWM, it is necessary to choose one from among the many candidate configurations, and the choice-RT RWM does not apply. In a prototypical choice RT task, the problem is to discriminate between two or more clearly above threshold stimuli. The moment of stimulus onset is obvious; the difficulty is in discriminating which of two or more possible stimuli occurred. The random walk in these cases is assumed to begin at the moment of stimulus onset.

*6.2.2.4. Random Walks for Go/No-Go Paradigms.* The go/no-go paradigm differs from the choice RT paradigms in that sampling of information cannot begin at a well-defined point of stimulus onset; it is stimulus onset itself that is to be decided. In the go/no-go case, especially with random foreperiods, the random walk must begin before the stimulus appears. This requires a random walk that begins after the warning signal and fluctuates around the bias point until stimulus onset. Once the stimulus appears, the parameters of the drift become stimulus dependent. On no-go trials, the stimulus never appears, and the observation interval typically ends when the experimenter terminates the trial. Thus, on signal trials, there are two phases to the random walk: (1) a period of fluctuation prior to stimulus onset with an expected value of zero, and (2) a subsequent period in which the drift is dependent on the identity of the stimulus. Analytic solutions for the predictions of an RWM where the characteristics of the drift alter in midtrial are not generally available (but see Ratcliff, 1980). Conceptually, the simple go/no-go walk can be considered to be equivalent to a stimulus-initiated random walk in which prestimulus fluctuations simply contribute to the intertrial variability in the effective starting point (e.g., $C_L$ in Figure 2.19(a)).

For a single location being monitored in a go/no-go RT experiment (the alone or baseline condition), consider an RWM with two boundaries. A near boundary is for the go responses, and another, much farther boundary corresponds to no-go responses. When the near go-boundary is reached by the random walk, a response is initiated. When the no-go boundary is reached, observation ceases on that trial, and preparation is made for the next trial. The no-go boundary has little influence on the simple go/no-go experiment: a distant no-go boundary guarantees there will be few trials where the stimulus appears, but the subject omits the response. The experimenter, not the no-go boundary, typically terminates processing on the catch (no-go) trials. The single-stimulus conditions are shown in Figures 2.19(a) and 2.19(b), with the distance to the go boundaries labeled $A_L(1)$ and $A_R(1)$.

When the subject is asked to monitor two locations simultaneously, this is modeled by two simultaneous go/no-go random

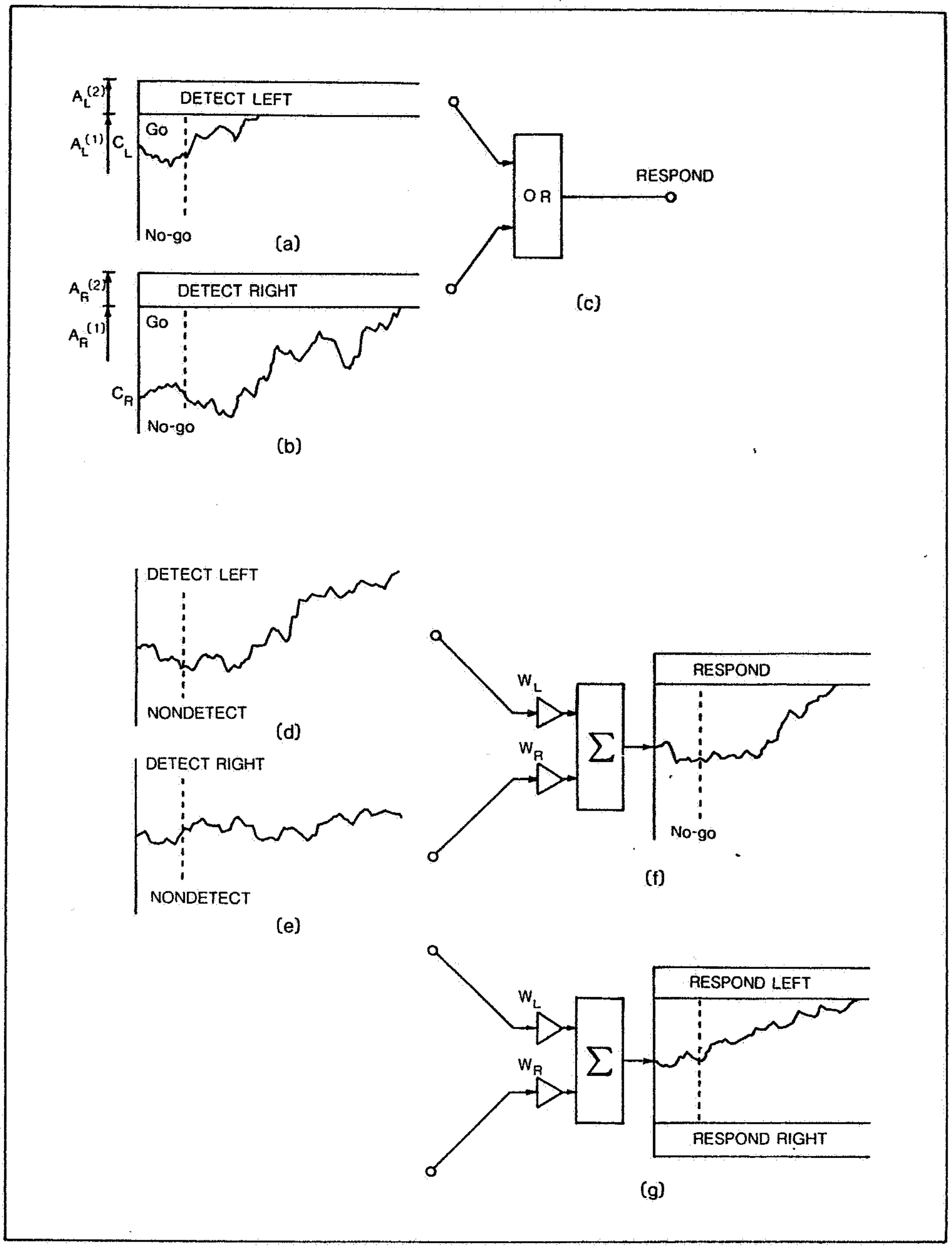



**Figure 2.19.** Random walk models (RWMs) for reaction-time (RT) experiments. (a) Representation of a RWM for a go/no-go RT experiment with a left stimulus. The coordinates are the value of the random walk and time $t$. The point $C_L$ on the left boundary represents the beginning of the trial; the vertical dashed line represents the onset of the stimulus on go trials. The random walk response boundary $A_L(i)$ is used for go/no-go trials on which there are $i$ possible stimulus alternatives. The no-go boundary (for "abandon trial") is not shown. (b) Representation of a RWM for a right stimulus with coordinates as in (a). Performance in a concurrent RT task (independent, concurrent presentation of tasks (a) and (b)) is modeled by the independent operation of random walk processes (a) and (b). (c) The additional apparatus needed to extend the single-stimulus go/no-go RT models to a multistimulus go/no-go task. Components (a) + (b) + (c) together represent a *pandemonium* (or *sensory threshold*) model. When either a left or a right random walk boundary is crossed, the OR component causes a predetermined response to be executed. The a priori probability $\alpha$ of a left stimulus determines the starting points $C_L$, $C_R$, shown here for $\alpha > 5$. (d, e, f) A *weighted decision* model for the multistimulus go/no-go task. The random walks (d) and (e), respectively, are equivalent to (a) and (b) except that (d) and (e) do not initiate responses. The current values of the walks in (d) and (e), respectively, are multiplied by positive weighting constants $W_R$, $W_L$, respectively, and summed, and the resultant walk initiates a predetermined response when it exceeds its threshold. A priori stimulus probabilities are reflected in constants $W_R$ and $W_L$. (g) A RWM for choice reaction time that differs for (d, e, f) only in that the stimulus walks are subtracted (right is multiplied by $-W_R$) and in that one of two alternative responses is admissible. (From G. Sperling, A unified theory of attention and signal detection, in R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

walks. Crossing *either* go boundary initiates a response. This is a *pandemonium* model in the sense that all sensory channels shout out their evidence simultaneously; if any one exceeds threshold, a response is initiated. (The pandemonium model is equivalent to the maximum rule of Section 4.2.3.) Obviously, two random walks would produce more false reactions to catch trials than one walk. Since subjects are instructed to avoid errors, they must move the boundaries farther away from the starting point to maintain the same accuracy in performance. This is illustrated in Figures 2.19(a) and 2.19(b) by increasing the distances to the boundaries, that is, by shifting from boundaries labeled $A_L(1)$ and $A_R(1)$ to boundaries $A_L(2)$ and $A_R(2)$. Because of the boundary shift, even in this RWM which has perfect retention of all information received from all channels, monitoring two locations produces slower decisions than monitoring just one.

Figures 2.19(d), 2.19(e), and 2.19(f) outline an alternative RWM for the same multistimulus go/no-go task. This weighted decisions model takes account of a priori stimulus probabilities in the weights $W_L$ and $W_R$ assigned to the sensory evidence being obtained from the left- and right-input channels. A response is initiated when the weighted, cumulated evidence exceeds a threshold.

*6.2.2.5. Compatible Models for Choice and for Go/No-Go Reaction-Time Tasks.* Both the pandemonium and the weighted decisions RWM models are easily elaborated to deal with choice RTs. The pandemonium model becomes a *race* model; the subject executes a left or right response according to whichever random walk reaches a boundary first. (That is, Figures 2.19(a) and 2.19(b) suffice to describe the race model.) The adaptation to choice RT of the weighted decisions model is illustrated in Figure 2.19(g); it merely involves changing the sign of one of the weights and admitting a second response alternative. An attractive feature of the weighted decisions models in Figures 2.19(f) and 2.19(g) is that they account nicely for the covariation, reported by Posner and colleagues (1978), between RTs in choice and in go/no-go tasks as stimulus probabilities are varied.

Both the race and the weighted decisions RWMs are different from Link's RWM (Section 6.1.3). The mathematical equivalences and differences between these various models have not been worked out. Deciding between such models, experimentally, requires complex paradigms and systematic data collection. Probably, different models will work best in different situations. The purpose of illustrating them here is to show a variety of models that have no internal attentional or memory losses and yet exhibit probability effects such as cost-benefits in RT paradigms.

**6.2.3. Random Walk Models and Signal Detection Theory: Bias versus Sensitivity.** The explanation of two-location compound RT tasks is exactly analogous to the explanation of the difficulty in searching for two targets ("1" or "2") instead of one target. The concurrent task of searching for ("1" and "2") does not have this problem. Nor would the concurrent task of presenting stimuli independently for responses with the left and right hands. The concurrent RT task is composed of two simple component RT tasks: respond with the left hand if left stimulus, respond with the right hand if right stimulus. The concurrent RT task, in which both left and right stimuli might occur on any given trial, is quite different from the usual disjunctive (choice) RT task, which is a compound time (RT) paradigm in which subjects responded directly to stimuli that were differentiated only by their location. Here we consider a choice RT experiment by M. L. Shaw (1978) in which spatial location and target identity are independently varied. Shaw's subjects were required to search arrays of $n$ locations for a single target (either $F$ or $Z$) among $n - 1$ distractors ($H, J, K, L, N$). On each trial the subject reported either "F" or "Z." The probability function, which gives the probability of the target appearing at each location, was held constant over a block of trials (instead of being cued on each trial). Shaw compared RTs for high-probability locations to RTs for low-probability locations. In Shaw's experiment, the location expectancy should not cause a bias to respond "*F*" or "*Z*," thus eliminating one of the explanations of the Posner and colleagues (1978) experiment discussed in Section 6.2.2.

M. L. Shaw fit her data with a quantitative attention model adapted from Koopman (1957). It is a model for optimal allocation of a limited search capacity and is discussed in detail in Section 7.3. The capacity-restricted search model, when applied to Shaw's experiment, explicitly assumes a limitation in attentional capacity and fits the data quite well.

Shaw's experiment, when all the complications are stripped away, is a choice RT experiment (respond "F" or "Z" as quickly as possible) and therefore is a *compound* task. The task requires a model for the stimulus uncertainty effect before an attentional loss can be inferred; an appropriate model for stimulus uncertainty in choice RT tasks is the RWM. An RWM for Shaw's data might involve a separate random walk between an $F$ and $Z$ at each location. The value of the random walk for each location might sum together, weighted by a fraction proportional to location probability. This is the random walk analog to the Nolte-Jaarsma optimal rule in SDT.

In the RWM, when a target appears at a low-probability location, the overall random walk to the $Z$ boundary is slowed because of the high weighting of information about distractors at high-probability locations. Such a scheme (see Figure 2.20) exhibits location-dependent reaction time, but this is due to differential *weighting* of information from different locations (a form of bias) and not due to any loss of information. The analogous model applied to detection paradigms has been called a *weighted decisions* model by Green and Swets (1966) and has been studied by M. L. Shaw (1982) and Kinchla and Collyer (1974). A detection paradigm was used by Bashinski and Bacharach (1980).

Both Shaw's model (based on resource allocation), and an elaborated RWM (based on weighted decisions) can fit Shaw's data reasonably well. This is another example of the ambiguity of compound tasks. The main question, Can attention be divided between locations without loss? cannot be answered without reference to a model. Further, the choice between the two contending models must be made on the basis of additional tests. The data may already be available. For example, both the elaborated RWM and especially Shaw's optimal-allocation model implicitly make strong predictions not merely about the RT means of correct and error responses but about the entire distribution of reaction times in the various conditions.

## 6.3. Speed–Accuracy Trade-offs

In RT tasks subjects traditionally are asked to respond as quickly as possible while making as few mistakes as possible. These are clearly incompatible goals; the subject could go faster by accepting more mistakes or could reduce errors by slowing down RTs. The ambiguity of the fast-and-accurate instruction

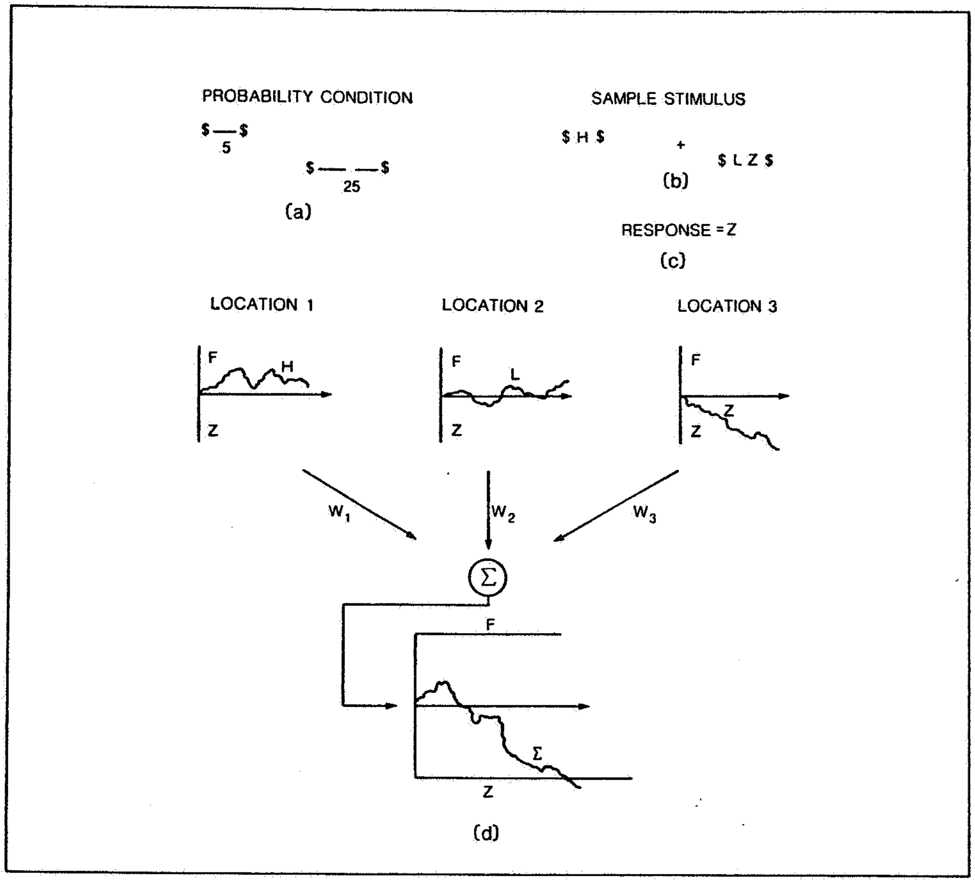



**Figure 2.20.** Optimization in letter search. (a) One probability condition from M. L. Shaw's (1978) letter search experiments. The subject's task is to identify the target, which is either an *F* or a *Z*. There are three spatial locations indicated by dashes; the number under a location indicates to the subject the probability that the target will appear in that location. (b) A sample stimulus display. The + indicates the fixation mark. (c) The response. (d) A *weighted decision* random walk model to account for the effect of location bias. The coordinates of each random walk (RW) are the value of the walk and time, *t*. A RW is carried out at each of the three locations, and the values are summed with a weighting, $W_i$, the location bias. The summed RW hits the appropriate boundary more rapidly when the target is in an expected location because it has a larger value of $W_i$. Shaw interpreted her results in the framework of Koopman's (1957) optimal search model, Figure 2.24.

is well known and, in better-designed contemporary experiments, the subject is rewarded according to a well-specified payoff matrix for quick correct responses and penalized for errors.

This section analyzes the implicit trade-off between speed and accuracy in the classical RT procedure, as well as in two variations. In the *deadline procedure*, the subject is given a time limit (the deadline) within which the response must fall to avoid an explicit penalty (Fitts, 1966). In the *cued-response procedure* (Dosher, 1976, 1981; Reed, 1973; Wickelgren, Corbett, & Dosher, 1980), a "respond-now" cue follows the stimulus with a variable delay. The subject is required to respond within a very brief interval (deadline) following the response cue.

To induce the subject to respond more quickly in the three procedures (classical reaction time, deadline, cued response), the rewards for fast and the penalties for slow responses are increased, the deadline is shortened, or the delay of the response cue is decreased. To induce the subject to be more accurate, the penalty for errors is increased, the response deadline is increased, or the delay of the respond-now cue is increased. Thus, given precisely the same stimuli, we can induce subjects to be either fast and inaccurate or to be slow and accurate. The range of performance of which a subject is capable defines the speed–accuracy trade-off.

**6.3.1. Deadline Speed–Accuracy Trade-off and an Analysis.** Figure 2.21 illustrates a typical speed–accuracy trade-off. The data are from a two-choice RT experiment by Pachella and Fisher (1972), with deadlines of 300, 400, 700 msec, and infinity (accuracy emphasis). To show the general form of the speed–accuracy trade-off, both accuracy and speed are averaged over all responses. Accuracy is presented as proportion correct, speed as the mean reaction time. To maintain the convention that good performance is represented up and to the right, fast RTs are represented to the right of slow RTs. Graphing the speed–accuracy trade-off in operating space emphasizes its similarity to other operating characteristics (attention operating characteristics, receiver operating characteristics, and production possibilities frontiers) and shows the relation of the speed–accuracy trade-off curve to optimization criteria embodied in utility functions.

Payoffs in speed–accuracy trade-off experiments usually are defined in terms of individual responses. Therefore, to compute the utility of a mean RT and a mean accuracy (averaged

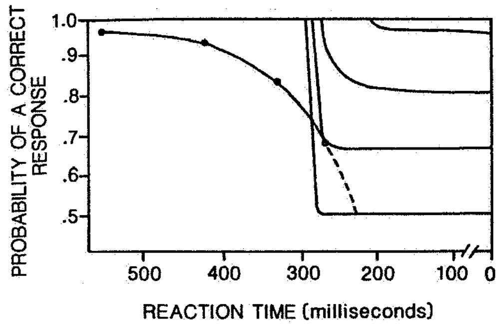



**Figure 2.21.** An operating characteristic and iso-utility contours for a speed–accuracy trade-off experiment using the deadline method. Data are derived from Pachella and Fisher's (1972) two-alternative reaction-time (RT) experiment. The abscissa is mean RT, with fast RTs at right; the ordinate is the probability of a correct response. Data are computed from Pachella and Fisher, (1972). Iso-utility contours are illustrated for a deadline of 300 msec, corresponding to the right-most data point. (From G. Sperling, A unified theory of attention and signal detection, in R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

over a session, as in Figure 2.21) requires knowledge about the actual distributions of RTs and error rates. The form of the utility function assumed here for deadline and cued response procedures is so simple, however, that it is relatively independent of the distributional details.

Figure 2.21 shows utility functions for a single trial with a 300-msec deadline in Pachella and Fisher's (1972) experiment. Although this experiment did not use an explicit payoff scheme, subjects were instructed to be as accurate as possible but *not* to exceed the response deadline. Utility is assumed to be proportional to the number of correct responses, with a very high penalty for responses that exceed the deadline, so the iso-utility function is vertical at the deadline. Responding much sooner than the deadline is not explicitly rewarded, so the iso-utility function is horizontal over all RTs shorter than the deadline, with higher accuracy having greater utility.

The experimentally defined utility for a single trial, illustrated in Figure 2.21, does not take into account irreducible variability in the subject's RT. For example, if the subject were to attempt to respond with an RT of 299 msec to beat a 300-msec deadline, then, because of RT variability, the RT would exceed the deadline on almost half the trials. To reduce post-deadline responses to a tolerable level, the subject has to aim response well in front of the deadline. Thus the utility function of strategies observed *over a whole session* must incorporate RT variability, and these utility functions are illustrated in Figure 2.21. Points in operating space have a utility corresponding jointly to the explicitly graphed accuracy (the ordinate of Figure 2.21) and to the fraction of postdeadline responses that are implicit in the graphed speed (the abscissa of Figure 2.21).

Quite generally, in nonpathological cases, the performance operating characteristic (POC) is concave down, the iso-utility contour is concave up, and the two curves are tangent to each other at the optimum point. In limiting cases, either the POC or the iso-utility contour may be straight lines. If both are straight lines, the POC and iso-utility contour may be co-linear, and the optimum performance is not uniquely defined. In the deadline and cued-response procedures, however, the "corners" of the iso-utility contours (where the tangent point of the speed–accuracy trade-off and utility function will be) tend to be almost vertically above each other (Figure 2.21), demonstrating the overriding importance of speed relative to accuracy in determining the operating point on the speed–accuracy trade-off.

How is it that subjects accomplish performance under a deadline procedure? One possibility, here designated as *information criteria*, is that an RWM applies, and the boundaries (information criteria) are set so that no more than, say, 5% of the responses exceed the deadline. Since most of the RT distribution must lie before the deadline, the mean RT would be to the right of the deadline. Alternatively, subjects could be *estimating a time interval* slightly shorter than the deadline and responding on the basis of whatever information is available at that time. The question of which strategy (information criterion or time estimation) is actually used is unresolved. Link (1978) has argued that information criteria (horizontal boundaries) adequately account for performance in deadline experiments. Wandell (1977), in a slightly different context, proposes that time estimation may be used under some circumstances.

**6.3.2. Cued-Response Speed–Accuracy Trade-off.** The purpose of the cued-response speed–accuracy trade-off procedure is to interrupt the subject's stimulus processing at some known time after stimulus presentation. By repeating the procedure with cues at different times, it is possible to determine the amount of processing that has been accomplished as a function of time following stimulus presentation.

In the cued-response speed–accuracy trade-off procedure, at some unpredictable time after stimulus presentation, the subject is cued with a secondary stimulus (tone or light flash) to respond immediately. Subjects are trained not to anticipate the cue, just as subjects in any RT experiment are trained not to respond in advance of the reaction stimulus. Subjects also are trained to respond as quickly as possible following the cue. Cue RTs should not exceed about 275 msec; ideally, mean RTs are under 225 msec, quite comparable to simple RTs to the cue stimulus.

The iso-utility contours in the cued-response procedure are steep U-shaped functions, intended to confine RTs to the interval defined by the *U*. The cued-response procedure constrains responses to a narrower time interval than does the deadline procedure as is obvious from the comparison of the utility functions in Figures 2.21 and 2.22. The deadline procedure is a blocked procedure; the cued-response procedure is a mixed-list procedure. (The implications of the blocked/randomized difference are treated in Section 8.)

A typical cued-response speed–accuracy trade-off function from Dosher (1984) is shown in Figure 2.22. Iso-utility contours are shown for a cue to respond 1.0 sec after stimulus presentation. As with the deadline procedure, the utility contours defined by the experimenter are simply rectangular (respond after the cue and before the deadline). As in Figure 2.21, the rounded shape in Figure 2.22 results from the subject's inability to control response latency perfectly. In principle, it would be advantageous for the subject (1) to aim safely inside the experimenter-defined boundaries and (2) to wait as long as possible to gain the most information. In practice, these options can be virtually removed by careful placement of the boundaries, that is, by the extreme pressure to respond to the cue as quickly as possible.

Information-controlled response strategies cannot account for the increase in accuracy with increasing cue delay in the cued-response paradigm. For example, to account for these increases in accuracy with cue delay, an RWM would have to have delay-dependent boundaries, that is, more distant boundaries for long delays. Because cue delays are randomly intermixed, subjects cannot make their processing strategy contingent upon cue delay (as they might if delays were run in blocks of the same delay). Therefore, the RWM cannot have delay-dependent boundaries, and therefore it cannot account for cued-response performance.

An alternative model for the cued-response procedure is an elaborated RWM with very distant (ignorable) informational boundaries. The subject's response is determined not by the random walk's intersection with a boundary, but by which side of its starting point the random walk happens to be at the moment the cue is presented (Dosher, 1982; Ratcliff, 1978). The RWMs for deadline and cued-response procedures are compared in Figure 2.23. The difference between models for these procedures is that performance in a deadline paradigm is accounted for by an *information-controlled* random walk (Figure 2.23(a)), while performance in a cued-response paradigm is accounted for by a *time-controlled* random walk (Figure 2.23(b)).

Subjects in the cued-response procedure must be trained to allow time-controlled processing. Early in training (in the first 100 trials), subjects have a strong tendency to anticipate the cue on late-cue trials and respond late to early-cue trials. Even with extensive practice, cue RTs (measured from cue onset) are somewhat longer for early cues than for late cues (Dosher, 1976, 1981, 1982; Reed, 1973). These RT patterns rarely depend on the stimulus; they resemble warning functions (Section 8).

Subjects in the deadline procedure also require practice to optimize their performance, especially to minimize the number of trials that exceed the deadline. Optimal setting of information criterion (horizontal) boundaries or time estimation (vertical) boundaries in an RWM requires sophisticated understanding of the RT distributions. Reaction-time distributions are generally more skewed and have a longer tail as mean RT (and accuracy) increase, and the subjects must take this into account in setting response boundaries.

**6.3.3. Iso-utility Contours from Reaction Times.** According to an optimization theory, even with the ordinary, ambiguous

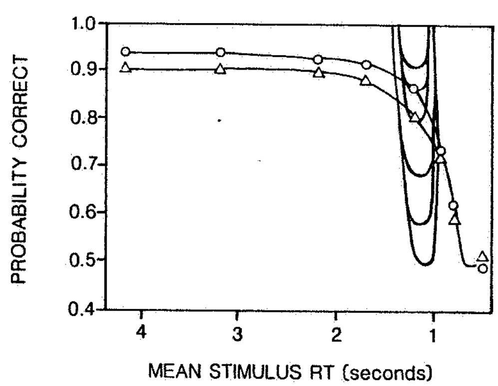



**Figure 2.22.** Operating characteristics and iso-utility contours for a speed–accuracy trade-off experiment using the cued-response method. The abscissa is the time (in seconds) from the stimulus onset to response (stimulus reaction time (RT)); the ordinate is the probability of a correct response. *Two speed–accuracy trade-off* curves (recognition accuracy for materials studied for 2 or 5 sec) observed under the cued-response procedure (Dosher, 1984) are shown, along with the hypothetical iso-utility contours for a cue to respond at 1.0 sec after test stimulus onset. Subjects are instructed *not* to anticipate the cue but to respond as quickly as possible—within 200 msec *after* the cue. The narrow windows defined by the utility curves of the cued-response procedure constrain the processing time very closely; they strongly penalize anticipations of the cue (fast guesses, cue RT < 100 msec) and slow responses (cue RT > 300 msec).

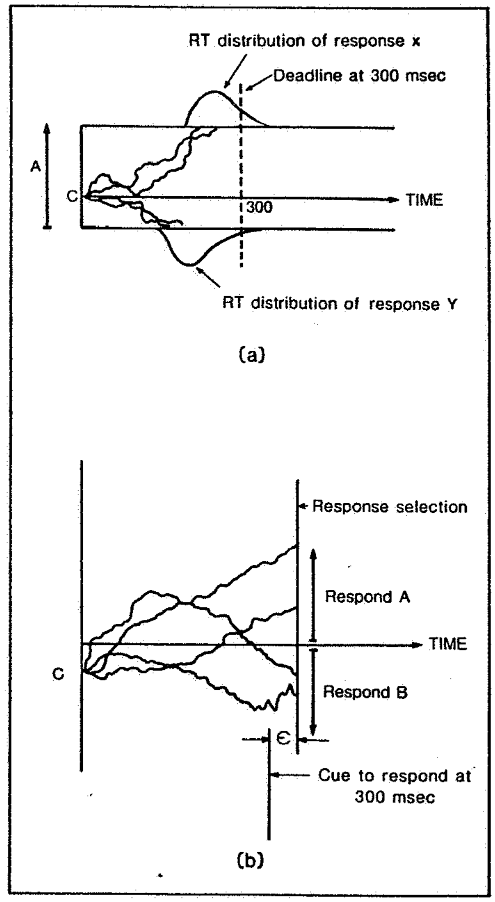



**Figure 2.23.** Random walk models (RWMs) of two speed–accuracy trade-off (SAT) procedures. For simplicity, in this figure, only decision processes (RWM) are assumed to contribute to overall reaction time. (a) A random walk (RW) interpretation of the *deadline* speed–accuracy trade-off procedure. The abscissa is time *t*, and the ordinate is the value of the RW at each time. Several representative RWs are shown. Stimulus *y* is much more probable than stimulus x. The RW boundaries (*A*) and the bias point (C) are set by the subject (after some experience) so that only a small fraction of trials produce RTs beyond the 300 msec deadline. Only RWM contributions to overall reaction time of (b) An RWM of the cued-response speed–accuracy trade-off procedure. A cue to respond is given 300 msec after stimulus onset; ε msec later, the subject selects the response. The boundaries *A* are absent, and the position of the RW relative to the ordinate at the time of response selection determines the response.

speed-plus-accuracy instruction, the subject operates at the optimal point on the speed–accuracy trade-off; with ambiguous instructions, the optimum is determined by the subject's *implicit* utility function. Insofar as different points on a speed–accuracy trade-off can be measured, the reasoning can be reversed and the tangent relation between the speed–accuracy trade-off and the iso-utility contour may be used to infer the shape of the subject's implicit iso-utility contours.

Finally, it should be noted that multiresponse RT experiments cannot be represented completely by a single speed–accuracy trade-off. For example, the speed and accuracy of particular response alternatives can be varied inversely even as overall performance averaged over alternative responses remains relatively unaffected, the problem to which RWMs are addressed. However, in symmetrical situations, where the difficulty and payoffs for the various alternative responses in the compound task are approximately equal, the speed–accuracy trade-off has interesting and useful properties (see Section 7).

## 7. PURE STRATEGIES

A subject's selection of a processing or decision strategy can have a substantial impact on task performance. Pure strategies are considered in this section, strategy mixtures in Section 8. First, optimum strategies for common psychological paradigms are given. Second, the dependence of the subject's strategy on the trial-by-trial structure of the experiment is illustrated with several detailed examples from classical procedures.

### 7.1. Definitions of Pure Strategy

**7.1.1. Unidimensional Strategy.** For the purposes of this chapter, a *strategy* is defined as the choice by which a subject arrives at a particular point along a particular performance operating characteristic (POC). The choice may involve several component processes, such as a decision criterion or a resource allocation decision. When only one independent choice is possible (e.g., decision criterion and resource allocation cannot be varied orthogonally), this is called a *simple*, or *unidimensional*, strategy. Varying the strategy parameter moves performance along a line in operating space, the POC. The dimensionality of the operating space itself is equal to the number of independent measures of performance.

**7.1.2. Multidimensional Strategy.** The definition of a strategy can be extended to multiple-independent choices. For example, when a two-dimensional strategy is varied appropriately and two or more responses are measured, the locus of accessible points is a POC surface (not a line) in multidimensional operating space. Strategy decisions may result in complex patterns on the POC surface. When the higher-dimensional data are projected into two-dimensional operating space, strategy variations (movements on the POC surface) can easily be confused with sensitivity variations (shifts to different POCs). The only multidimensional strategies considered in this chapter occur in reaction-time (RT) models (such as the random walk model) where the subject can choose a different strategy for each response alternative.

Strategy is a real-world term; that is, it refers to something the subject does or is assumed to do. In a model or in theory of performance, a strategy is represented by a decision rule and a single parameter; different strategies are represented by different values of the parameter. To avoid the cumbersome "the subject selected the strategy $S_c$, which is represented in the model by the decision rule with parameter $c$," one may say simply "the subject selected $c$," keeping in mind that data analysts, not subjects, select parameters. In the classroom model, a strategy $S_c$ is represented by the choice of classroom switching time $c$. In signal detection theory (SDT) a strategy $S_c$ is defined as the value $c$ "chosen" for the decision criterion. In attention or economic theory, a strategy is represented by the parameter that describes the resource allocation choice. In RT tasks, a strategy is the selection of a joint level of speed and accuracy along the speed–accuracy trade-off for each of the response alternatives. In the random walk model, this multidimensional strategy is represented by the values of the response boundaries.

### 7.2. Optimal Strategies for Psychophysical Paradigms with Two-Stimulus Alternatives

**7.2.1. Single-Stimulus Presentation (Yes-No).** Signal Detection Theory, unelaborated, deals with the problem of an optimal strategy in the case of a single-stimulus presentation. The stimulus may be either $N$, or $S + N$, and the subject responds either "$N$" or "$S$." In this chapter, the decision rule that has been proposed is the choice of a criterion $c$ on a psychological continuum. That is, from the subjects' point of view, the outcome of observing the stimulus on a trial is the perceived stimulus intensity, which is represented by a real number $x$. If $x \geq c$ the subject responds "$S$," otherwise, "$N$." In the case of discriminating stimulus $A$ from stimulus $B$, the argument is exactly parallel; the continuum represents the ratio of perceived "$A$-ness" to "$B$-ness."

The optimum decision rule is correctly expressed not in terms of a sensory variable but in terms of the *a posteriori* likelihood ratio, $aplr = p(S|x)/p(N|x)$, where $p(E|x)$, the likelihood of $E$ given $x$, is the *a posteriori* probability of the event $E$ having occurred given the observation $x$. When the utility of a trial is defined in terms of the value of various outcomes of a trial, decision rules that maximize *expected utility* are based on the *a posteriori* likelihood ratio. For example, the optimal decision rule is of the form: respond "$S$" if $aplr(x) \geq c$, otherwise respond "$N$."

The formulation of optimal decisions in terms of *a posteriori* probabilities is conceptually basic. For computation, the equivalent formulation in terms of *a priori* probabilities is more useful. Thus, if $\pi_i$ is the *a priori* probability of the event $i$, the optimal decision rule transposes to: Respond '$S$' if the likelihood ratio $lr(x) = p(x/S)/p(x/N) \geq \pi_N/\pi_S$. Recall that the decision rule based on a sensory variable is equivalent to the likelihood rule when $lr(x)$ is an increasing monotonic function of $x$ (as it was in the examples of this chapter); otherwise, only the $lr$ rule is optimal. Further, the likelihood-based rule is optimal under much broader definitions of utility, but these considerations are beyond the scope of this chapter.

**7.2.2. Two-Stimulus Presentations**

***7.2.2.1. Two-Alternative Forced Choice.*** In the two-alternative forced-choice (2AFC) procedure, a subject is presented with two stimuli, $A,B$ (which represent, for example, $N$, $S + N$). The stimuli may occur in successive intervals, or they may occur in adjacent locations. The subject's task is to state whether the order of presentation was $AB$ or $BA$. In the case of adjacent presentations, the subject's task is to state whether the target occurred in the left location or in the right location. This procedure grew out of signal detection considerations because when rewards for both kinds of correct responses are equal and penalties of both kinds of errors are equal (symmetric payoffs), it appears to remove the choice of a decision criterion from the task. The optimal decision strategy is simply to compare the observations $x_1$, $x_2$ from the two intervals and to report the order as $AB$ if $aplr_{A|B} = p(A|x_1)/p(A|x_2) \geq 1$ and report the order $BA$ otherwise.

For the special case of Normally distributed random variables with equal variance, the optimal decision (in asymmetric as well as symmetric cases) can be made on the basis simply of $x_1 - x_2$. Indeed, in the Normal case, the decision variable $(x_1 - x_2)/\sqrt{2}$ in two-alternative forced choice is equivalent to the decision variable $x$ of the yes-no procedure. The discussion here is restricted to symmetric payoffs and equal probabilities of the signal occurring in each interval. For additional assumptions, for treatment of asymmetric situations, and other complexities, see Noreen (1981).

***7.2.2.2. Same-Different Paradigm.*** As in two-interval forced choice, there are two stimulus presentations, $A,B$, but the correct responses are "same" for presentations of $AA$ or $BB$ and "different" for $AB$ or $BA$. The optimal decision strategy (derived by Noreen, 1981) requires the subject to make two separate categorizations of the two-stimulus observations $x_1,x_2$. The procedure is (1) first categorize $x_1$ as $A$ if $aplr_A(x_1) = p(A|x_1)/p(B|x_1) \geq 1$ and otherwise categorize $x_1$ as $B$, (2) then categorize $x_2$, and (3) if $x_1$ and $x_2$ have been categorized as $A,A$ or $B,B$ respond "same," otherwise respond "different." Self-evident as the categorize-first rule may be, a sensory different rule (respond "same" if $|x_1 - x_2| < c$, otherwise respond "different") was previously described as the optimal rule (Krueger, 1978; Macmillan, Kaplan, & Creelman, 1977; Vickers, 1979). The advantage of the categorize-first rule is that two events, both of which are very likely to be $A$, may still differ greatly and be categorized optimally; the sensory-difference rule would miscategorize them. The advantage of the sensory-difference rule is that it is applicable to the case of the "roving standard" or to early stages of practice before categorization boundaries have developed. (The difference rule yields equivalent statistical predictions to the optimal rule in the special case of two-alternative forced choice under the assumption of equal-variance Normal distributions.)

The actual differences that would be observed between the optimal early-categorization and the sensory-difference rules are slight in practical situations. The great interest that the analysis of the same-different paradigm has aroused is due to its incorrect application of sensory-difference rules to the problem of categorical perception (Macmillan et al., 1977). According to the optimal-decision model, same-different discrimination *is* essentially a process of categorization, even though, at first glance, it seems ideally suited for a sensory-difference strategy. Since the predictions of the two decision rules differ slightly, and in some cases not at all, the same-different paradigm does not seem to be an arena in which the issue of categorical versus sensory-difference rules will be decided.

#### 7.2.3. Three-Stimulus Presentations

*7.2.3.1. ABX.* In the most common variant of $ABX$ three-stimulus paradigms, there are only two alternative stimuli $A,B$. A trial consists of three-stimulus presentations: $ABX$ or $BAX$, where $X = A$ or $B$. The subject's task is to state whether the third stimulus is the same as the first ("first") or the same as the second ("second").

According to the decision model, the subject extracts three observations $x_1,x_2,x_3$ from a trial. The optimum strategy (derived by Noreen, 1981) is an extension of the optimal rules for two presentations. It is (1) categorize $x_1$ and $x_2$ as either $AB$ or $BA$ (exactly as in the two-alternative forced-choice paradigm), (2) categorize $x_3$ as $A$ or $B$ (as in the yes-no paradigm), and (3) respond "first" if the resulting categorizations are $ABA$ or $BAB$, otherwise, respond "second." A sensory-difference strategy in which $|x_3 - x_1|$ and $|x_3 - x_2|$ are compared is almost optimal in typical situations.

*7.2.3.2. AXA.* In this paradigm, the subject must decide whether the middle one of three presentations is the same or different from the outer two. The possible presentations are $AAA, BBB$ ("same") and $ABA, BAB$ ("different"). Upon decision analysis, this task becomes equivalent to $ABX$—the two identical stimuli ($AA$) supplying exactly the same benefit as the two complementary stimuli ($AB$) in $ABX$. For analyses of still other paradigms, see Noreen (1981).

**7.2.4. Conclusions.** The conclusion about optimum strategies in the most common paradigms with multiple presentations are:

1. Compute the likelihood of each stimulus alternative $(A,B)$ for every observation $x_i$.
2. In two-alternative forced choice, use the likelihood ratio to decide whether the first- or second-stimulus presentation is more likely to have been $A$. In the other paradigms, categorize each of the observations as the more likely of $A$ or $B$.
3. Formulate the appropriate response based on the categorizations.
4. An alternative approach is forming *sensory differences*. (e.g., $|x_i - x_j|$) and making decisions based on these. A strategy based on sensory differences is exactly equivalent to the *early categorization* strategy in some paradigms with some distributional assumptions (two-alternative forced-choice, equal-variance Normal distributions) and only slightly inferior in others of the paradigms considered here. The sensory-difference model is nonoptimal because it depends only on *relative* information (how different is $x_1$ from $x_2$) and neglects the absolute information that controls the independent categorization of the $x_i$ in the optimal model. The two models differ in their responses when (1) the sensory difference is small but the two observations fall on opposite sides of a criterion or (2) the sensory difference is large but the two observations fall on the same side of a criterion value. Since the difference model and the optimal model would result in the same response on most trials, the quantitative predictions differ only slightly.
5. As in the analysis of $n$-state threshold versus continuous theories in signal detection, and in Nolte and Jaarsma's (1967) analysis of ideal detection in multichannel listening experiments (Figure 2.10), two apparently quite different strategies can lead to very similar performances. Evidence for the use of one versus another of the decision strategies usually requires consistent data from more than one paradigm.

## 7.3. Resource Allocation: Optimal-Search Strategy

The concept of an *optimal* performance probably is best known in psychology in the context of SDT. It involves the question of whether subjects can perform (set a criterion in the signal detection model) according to an optimum-decision rule, for example, a maximum-likelihood rule. The notion of optimization (the maximization of utility) applies not only to decision strategies for dealing with incomplete information but also to resource allocation decisions. This section describes an important theory of the optimal *allocation* of search resources.

In the 1950s, Koopman (1956a, 1956b, 1957), the mathematician, derived a theory of search that defined the optimal allocation or distribution of limited resources for searching for the location of a target. This theory was originally developed in the context of military applications, for example, aircraft searching optimally for a submarine, given a limited number of flying hours. The basic assumptions are (1) a fixed, limited search capacity, (2) no cost for sharing search capacity among several locations and no cost for changing the allocation strategy, (3) a known probability density for the target in each location, and (4) a principle of diminishing returns in search efficacy. Koopman's theory can be applied only to compound tasks, since it assumes that the target is in *one* particular location.

Koopman's theory provides an algorithm for defining the amount of search effort that should be allocated to different locations under an optimal-allocation policy. In one dimension, the theory deals with a target located somewhere on a line, the $x$-axis, on a graph as in Figure 2.24. The $y$-axis represents the natural logarithm of the probability density function ($\ln p(x)$)

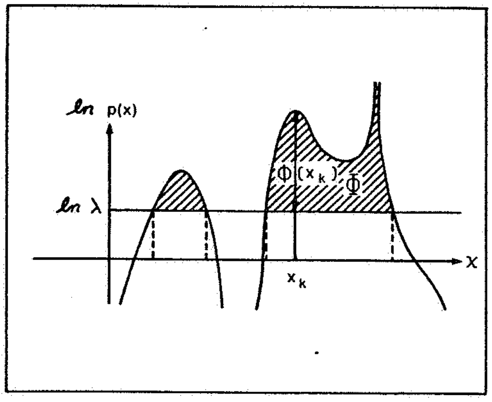



**Figure 2.24.** The graphic interpretation of Koopman's (1957) solution for determining the optimal distribution of search effort for detecting a single target whose location is unknown. The algorithm assumes that (1) the *probability* $p(x)$ of the target being at any location $x$ is known, (2) there is a limited pool of search effort, and (3) the probability $f(x)$ of detecting a target at location $x_k$ is $f(x_K) = 1 - e^{-\varphi(x_K)}$, an exponential (diminishing returns) function of the search effort $\varphi(x_k)$ allocated to $x_k$. The graph illustrates the special case where the target appears somewhere on a line. The abscissa represents location $x$; the ordinate is $\ln p(x)$. To determine the optimal distribution over $x$ of search effort, $\varphi(x)$, first graph $\ln p(x)$. Then draw a horizontal line such that the area between $\ln p(x)$ and the line (shaded area in the graph) is equal to the available search capacity $\phi$; this line is labeled $y = \ln\lambda$. Optimal allocation of search effort requires $\varphi(x) = 0$ for all positions where $\ln p(x) < \ln\lambda$, and $\varphi(x) = \ln p(x) - \ln\lambda$ for the remaining (high probability) locations. The area between $p(x)$ and $\lambda$ on a linear graph (i.e., $\int p(x) - \lambda$, for all $x$ where $p(x) > u/\lambda$) is the probability that the search is successful. (From B. O. Koopman, The theory of search III: The optimum distribution of searching effort, *Operations Research*, 1967, *5*. Reprinted with permission.)

for the target on the axis. To compute the optimal search strategy, the following graphic construction is used. A line parallel to the axis is adjusted in height ($\ln\lambda$) such that the area above the line and below $\ln p(x)$ equals a constant that represents the fixed capacity $\Phi$. Then only those locations where the function $\ln p(x) > \ln\lambda$ are searched, and the amount of effort allocated to the searched locations, $\varphi(x)$, is $\ln p(x) - \ln\lambda$. This is shown graphically in Figure 2.24.

One useful aspect of Koopman's (1957) model is that if more capacity or resources ($\Phi'$) become available after some search has been completed on the basis of the original resource commitment ($\Phi$), the algorithm can simply be applied on the a posteriori probabilities. Successive applications of this sort yield the same optimal distribution of resources as if ($\Phi'$) had been known originally.

Koopman's theory of optimal-detection search was imported into psychology by M. L. Shaw (1978; M. L. Shaw & P. Shaw, 1977) and applied to the problem of visual search for a target among distractors where the probability of a target at several locations is known. M. L. Shaw and P. Shaw (1977) applied this model directly to an experiment with detection probability as the dependent measure, and M. L. Shaw (1978) extended it to handle RT data. The issue was whether her subjects demonstrated an optimal distribution of attentional capacity. A related model for *optimal localization*, or "whereabouts," search was developed by Tognetti, 1968, Kadane, 1971, and Stone, 1975, and was extended and applied to visual search data by M. L. Shaw, Mulligan, and Stone, 1983.

In a search experiment with just two target locations, one likely target location and one unlikely location, Shaw's adaptation of the optimal-allocation model to RT data uses Koopman's principle of reallocation to predict that subjects should allocate all their search capacity to the higher-probability location until its posterior probability equals the a priori probability of the low-probability location (M. L. Shaw, 1978). Subsequently, the subject should divide capacity equally. The posterior probability of any location diminishes with search capacity already spent because a target in that location would likely have been detected given a reasonable expenditure of search resources. Shaw's data were fit reasonably by this model. The application of Koopman's model to RTs yields an additional strong prediction that apparently was overlooked by Shaw. The RTs for detections at the high-probability locations should be a mixture of two distributions resulting from the two levels of attention allocation and hence should obey the fixed-point property (cf. Falmagne, 1968).

From the psychological point of view, one interesting assumption of Koopman's model is that of no changeover or sharing costs. The "optimal" search allocation would differ from this if it were necessary to include changeover cost in the computations. In particular, changeover cost should lead to perseveration in a given allocation scheme past the point where the a posteriori probabilities would indicate an alteration of allocation. Changeover costs in visual search are considered in Section 9.

## 7.4. Strategies: Mixed-List and Blocked Designs

Many situational factors may influence a subject's selection of a particular strategy. Previous sections have considered the importance of task instructions, a priori stimulus probabilities, and payoffs in determining attentional allocation, decision criterion, and the speed-versus-accuracy of responding. Another factor, also under experimenter control, is the kind of stimulus mixture: which stimuli are presented in separate blocks and which stimuli are presented within a block. Such experimenter decisions can have a large impact on the strategy employed by the subject and on the interpretation of the data by the experimenter.

A *blocked* experimental design is one in which a set of experimental parameters $\Omega$ is kept constant for a block of trials (usually 100 or more) and $\Omega$ is varied between blocks. For example, $\Omega$ may represent the intensity of a tonal signal in a signal–noise detection task. The corresponding *mixed-list* experimental design (mixed blocks) is one in which all the values of the parameters $\Omega$ can occur within any block of trials. In the mixed-list experiment, all the types of trials that have been segregated into separate blocks in the block design are mixed together and presented in random order.

In terms of the compound-concurrent analysis of experimental procedures, the mixed-list procedure is a compound task; the individual blocks of the blocked procedure are the component tasks. Section 4 compared component to compound tasks in terms of *stimulus uncertainty* and outlined procedures for drawing correct inferences from experiments that varied stimulus uncertainty between conditions. This section continues the development of compound tasks in terms of the specific strategies adopted by the subjects.

Practical considerations influence the choice of experimental design. Historically, with manually operated apparatus, it was impractical to vary conditions between trials, so many experiments were run in block designs, not by choice but by necessity. To offset learning, fatigue, and other extraneous changes between blocks, complex, counterbalanced experimental designs were required. These were inefficient procedures because they required a commitment in advance to a particular number of subjects (required for counterbalancing) and an experiment of a particular size (that may have been too small or too large).

With the advent of computer-controlled experimentation in which new stimuli are generated cheaply and quickly, mixed-list designs became not only practical but also preferable. The conditions are mixed together, so the counterbalancing problem is instantly solved. A few subjects can be run for large numbers of trials. If data from a session are lost, it does not spoil the experimental design. Data collection is continued until a stopping criterion (such as a certain level of statistical reliability of estimated parameters) is reached; the amount of data may be different for different subjects. Such sequential statistical procedures are far more efficient than fixed procedures—an instance of optimizing the experimenter's strategy.

To illustrate the often crucial importance of the choice of mixed-list versus block design procedure, two kinds of commonly used paradigms are considered. In the first (exemplified by partial report, Section 7.4.1), a mixed-list design is essential for interpretation, but a block design is usually used; in the other (the method of constant stimuli, Section 7.4.2), the mixed-list design is universally advocated but leads to data that are awkward to interpret whenever SDT is appropriate.

**7.4.1. Pure versus Mixed Blocks in Information Processing Experiments.** Consider experiments with briefly presented stimuli in which presentation time is a critical parameter. For example, a row of letters is briefly exposed for a duration $D$. In a recall task, the subject must report as many letters as possible. In a search task, the subject may be required to say whether the stimulus contains the letter $q$. In these tasks, the experimenter wishes to determine the level of performance as a function of exposure duration. Should the various durations under investigation be run in separate blocks or together in a mixed list? The answer depends on the purpose of the experiment.

***7.4.1.1. Equivalent Processing Assumption.*** Whenever stimulus duration is varied, the *equivalent processing assumption* almost invariably is made implicitly. To illustrate: suppose the following stimulus durations are being experimentally tested: 50, 100, 200, and 500 msec. The theorist assumes that during the first 50 msec of the 100-, 200-, and 500-msec exposures, the subject processes information exactly as in the 50-msec exposure, and only after that time does processing differ. The equivalent processing assumption is especially important in comparing the longer durations, 200 and 500 msec, since they overlap much more.

The equivalent processing assumption is valid only in the mixed-list design. In the block design, the subject may (and usually does) employ different strategies in different blocks. For example, the subject may attend to the center of the display in brief exposures but attempt to process the display from left to right in longer exposures. In very brief exposures, *exposure* duration effectively controls apparent contrast but is ineffective in controlling *processing* duration (because of visual persistence). Since strategies tend not to vary enormously with contrast, the misinterpretations of exposure duration experiments did not become serious until the introduction of postexposure visual noise fields (Sperling, 1963) to interrupt processing. An instructive bad example in which the equivalent processing assumption is made incorrectly is described in Section 7.4.1.2.

***7.4.1.2. Equivalent Processing in Whole Reports.*** Sperling (1967) exposed a row of five letters for various durations, followed by a noise field, and determined the rate at which his subjects acquired information from each of the five letter locations. However, the experiment was run in a block design, so the assumption that the subject viewing long exposures was doing the same initial processing as at short exposures was unwarranted. Such an assumption requires a mixed-list design. In a block design, it is not even necessary that performance increase monotonically with exposure duration at each of the various locations. For example, in a long-exposure duration block, a subject may neglect a location $j$ (in favor of attending to location $k$) although reporting location $j$ accurately at short-exposure durations. This kind of paradoxical nonmonotonicity is ruled out in mixed lists.

***7.4.1.3. Equivalent Processing in Partial Reports.*** Partial report experiments are a particular trouble area for block designs. In the partial report procedure, a subject is presented with more stimulus information than can be recalled, for example, a brief flash of a $3 \times 3$ array of letters. The subject is required to report only one of the three rows. The cue that informs the subject which particular row is required (e.g., a high-, medium-, or low-pitched tone) occurs only after a delay of $D$ msec after the stimulus has been turned off. The logic of this experiment is that the subject cannot report all the letters (because of a recall-memory limitation) but can nevertheless give perfectly accurate partial reports as long as the stimulus is stored in a visual sensory memory. When the cue is delayed, the contents of visual sensory memory have decayed, and partial reports are less accurate. The decay of partial report performance with increased cue delay is assumed to represent the decay of sensory memory with the passage of time.

Partial report is a procedure crying for the delays to be run in mixed lists, yet they are nearly always run in blocks. It is simply wrong to assume that a subject waiting for a cue in a block of 500-msec delays is as passive during the first 150 msec as in a block of 150-msec cues. In blocks of short delays the subject may wait passively for the cue (equal-attention strategy), whereas in long-delay blocks the subject may begin to encode a particular row for response as quickly as possible following stimulus onset (a strategy of guessing which row will be cued). This block-dependent strategy is so obvious that it was described in the original partial-report study. Figure 2.25 exhibits data from a subject who failed to switch soon enough between an equal-attention strategy and a guessing strategy in successive blocks of trials, as the cue delay gradually increased (open circles) or gradually decreased (closed circles) between blocks. The data exhibit classical *Einstellung*, or, as it is often called now, *hysteresis*, where performance on one block of trials depends on the strategy chosen in previous blocks. Without the knowledge or control of the subject's strategy, it is not possible to estimate either the capacity or the duration of sensory storage. For example, the fact that Sperling's subject's performance reached asymptotic accuracy at cue delays of 0.5 sec may mean that the subject switched strategy to guessing in that block, not that the sensory store was empty.

***7.4.1.4. Conclusion.*** The important lesson is that the particular sequence in which trials are presented has an enormous influence on the subject's strategy and, thereby, on responses. In a block design, the subject can choose an optimal strategy for each block. In a mixed list, the subject must choose one strategy (or one mixture of strategies) for all the trials in the list. To be sure that a subject uses the same strategy in different conditions, these conditions must be run together in a mixed list.

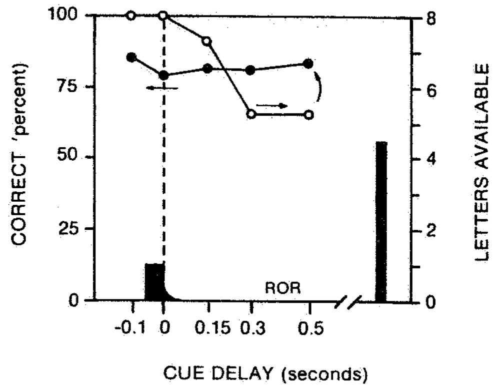



**Figure 2.25.** Partial report accuracy as a function of cue delay: an example of path dependence (hysteresis) in the selection of attention strategies. Cue delay was varied between 10 blocks of trials in 1 session; the arrows indicate the order of blocks, beginning with the series of increasing delays. The bar at the right shows the accuracy of whole reports. Increasing cue-delay blocks (open circles) correspond to a strategy of equal attention to the two rows of letters in the stimulus; decreasing cue-delay blocks (closed circles) correspond to a strategy of attending primarily to the top row. (Subject ROR from Sperling (1960), Fig. 5.) (From G. Sperling, The information available in brief visual presentations, *Psychological Monographs*, 1960, *74* (11, Whole No. 498). © 1960 by the American Psychological Association. Reprinted by permission.)

**7.4.2. Pure versus Mixed Lists in Detection: Method of Constant Stimuli.** Consider an experiment in which the subject must detect the presence of a tonal stimulus in noise and the tone occurs at several intensities within a block of trials. Note that the noise stimulus alone (blank, catch trial) must occur at least occasionally within the block; otherwise, the subject could simply report "signal" on all trials without observing the stimulus. Since several stimulus intensities appear randomly within a block, this is a compound task. The probability of a "signal" response as a function of signal intensity is called the *psychometric function* (Figure 2.26). The *method of constant stimuli,* just described, traditionally has been used to yield the psychometric function (Woodworth & Schlosberg, 1954).

The tone-in-noise discrimination task is well described by SDT. The difficulty detecting a tone in noise comes about because the noise contains energy at and near the tonal frequency. For many kinds of noise, including white noise, the amount of this energy is Normally distributed from trial to trial, with mean and variance, $\mu_N$, $\sigma_N^2$. On signal trials, the tonal energy is Normally distributed with mean and variance $\mu_N + \mu_S$, $\sigma_N^2$. An ideal detector, confronted with the signal-in-noise discrimination task, will compute these energy statistics and use them in exact conformity with psychological SDT. A human observer need not necessarily operate in conformity with SDT, but in situations like tone-in-noise discrimination in which the assumptions of SDT theory are virtually built into the physical stimuli themselves, humans conform quite closely to SDT.

Unfortunately, the method of constant stimuli does not fare well under analysis by SDT. The utter futility of the classical procedure of absolutely prohibiting false alarms (positive responses on noise-alone trials) and the enormous dependence of the psychometric function on the penalty for false alarms have long ago been extensively documented (Green & Swets, 1966) and need not be detailed further here. Payoff manipulations move the psychometric function to the left or right (as it is usually plotted) but leave the monotonicity properties intact; that is, the various psychometric functions obtained with different payoffs do not cross. (For an elaborate analysis of noncrossing psychometric functions, see Kruskal, 1965; Levine, 1971.) The effect of varying trial mixtures is more serious. According to SDT the various psychometric functions generated by block designs do cross (are not monotonically related, do not lie uniformly above or uniformly below) the psychometric functions generated in mixed-list designs. In the following, nonmonotonicity is demonstrated, and various remedies for the lack of a unique or generic psychometric function are considered.

The $N$ and the $S + N$ probability density functions (*pdfs*) that are at the core of the SDT analysis of the tone-in-noise experiment are illustrated in Figure 2.26(a). In the method of constant stimuli, the task is discrimination of noise alone from the compound alternative, the union of the various signal stimuli, which is represented in Figure 2.26(a) by the sum of the various signal density functions ($S_1$, ..., $S_6$). The decision criterion is set in accordance with expected utility which depends on the payoff matrix. Even in the traditional method of constant stimuli, in which the instruction is to avoid false alarms absolutely while detecting as many stimuli as possible, the payoffs cannot be defined simply in terms of penalties for false alarms, or the subject would never observe the stimulus but would merely respond "noise" on all trials. Once payoffs are defined, SDT applies and an optimum criterion can be selected. For the usual case of equally likely signal stimuli and the case of symmetrical payoffs for correct and incorrect responses, the optimum criterion ($c_1$) occurs where the $N$ *pdf* crosses the compound $\Sigma S_i$ *pdf* in Figure 2.26(a). The psychometric function generated by the SDT parameters corresponding to these experimental conditions is illustrated in Figure 2.26(b). It is generated by considering where each individual stimulus *pdf* lies with respect to the criterion. Provided that the payoff matrix retains positive payoffs for correct and negative payoffs for incorrect responses, changing the payoff matrix only moves the psychometric function to the left or right in Figure 2.26(b) but does not alter its shape. Figures 2.26(a) and 2.26(b) also illustrate the case where the penalty for false alarms (and the reward for correct rejections) is 6 times greater than the penalty for misses and the reward for correct detections. This case of 6 to 1 *payoff ratio* is represented graphically in the same way as a case of equal payoffs in which the *frequency* of noise stimuli is increased sixfold.

The method of constant stimuli can be analyzed as a compound task (or mixed-list experiment) in which the corresponding component tasks (or pure blocks) each contain noise and just one of the stimuli. Let the trials of the mixed block (which contains the noise and six stimuli) be separated out into six pure blocks, each of which contains just one signal intensity and just one-sixth of the noise trials. The pure block experiments are represented in Figure 2.26(c) corresponding to the two different payoffs (or to two different a priori stimulus probabilities) as described for Figure 2.26(a). Given the same payoff matrix as in the method of constant stimuli, the decision criterion differs for each of the pure blocks, and the optimum criterion occurs at the point where the noise *pdf* crosses the signal *pdf*. The corresponding psychometric function for the pure block experiments, curve $p$ in Figure 2.26(d), is flatter than and crosses the function $m$ for mixed blocks. However, a psychometric function is not really appropriate for pure blocks because the criterion varies between blocks. An accuracy function, such as percent correct (from which $d'$ is computable), describes the data better. The overall expected accuracy in the mixed blocks is, inevitably, somewhat lower than in the pure blocks, Figure 2.26(e), illustrating once again that signal uncertainty causes a performance

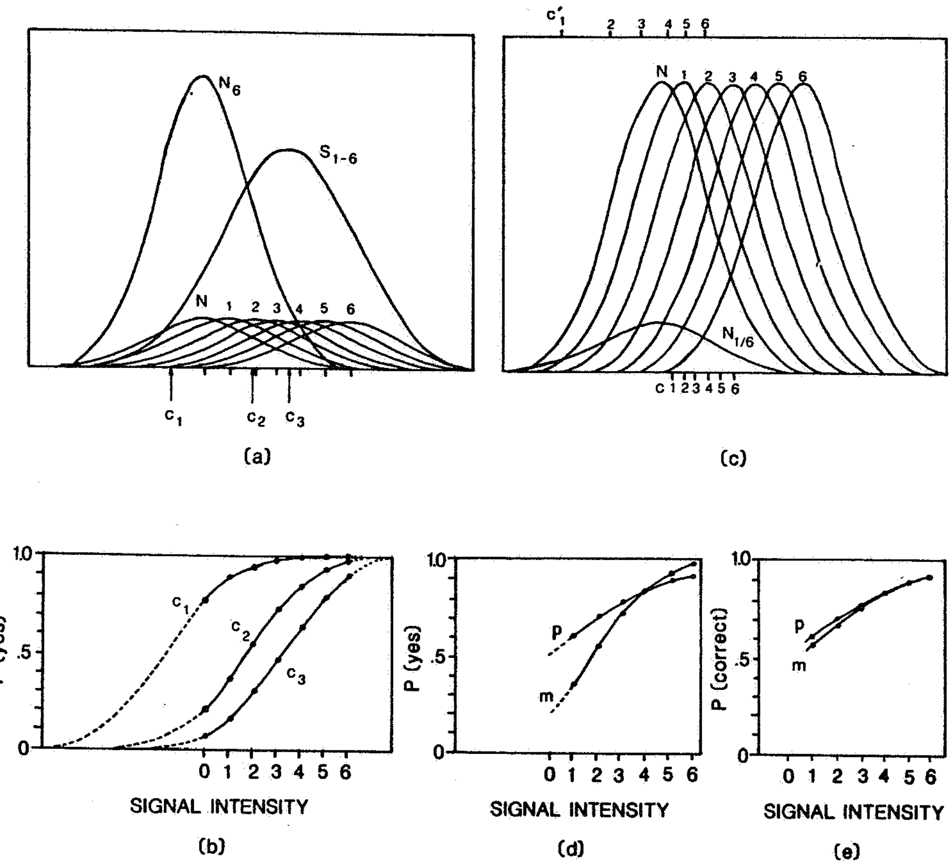



**Figure 2.26.** Analysis of the method of constant stimuli and the corresponding blocked designs. (a) Probability density functions (*pdfs*) of the assumed internal effects of the stimulus (noise-alone labeled *N*) and six different signal-plus-noise stimuli (signal labeled 1, ..., 6 in order of increasing signal intensity). $S_{1-6}$ is the sum of the six signal distributions and represents the compound *pdf* of the stimulus for correct "yes" responses in the method of constant stimuli. $N_6$ represents the *pdf* for noise multiplied by a factor of 6 to represent either a penalty 6 times greater for incorrect responses on noise than on signal trials or a sixfold increase in the frequency of noise stimuli. The point $c_2$ represents the criterion that maximizes expected utility in the discrimination of $N_6$ versus $S_{1-6}$. Criterion $c_1$ maximizes expected utility in the discrimination of *N* versus $S_{1-6}$—the case of equal probability of occurrence of each stimulus (noise and the six signals) and equal payoffs for correct and incorrect responses to each stimuli. Criterion $c_3$ restricts errors (false alarms) to exactly 5% when the noise stimulus is presented. (b) Probability of a detection ("yes") response as a function of signal intensity, where zero represents noise-alone. The predicted *p*("yes") for each of the three criteria in (a) are indicated by filled circles connected by heavy lines. (c) Optimal criteria in block designs. The *pdfs* (*N*,1, ..., 6) represent noise and the six signals as in (a). In a block composed of equal numbers of *N* and signal *i* ($i = 1, ..., 6$) trials, with symmetric payoffs, the optimal criterion $c_i$ occurs where the *N* and signal-*i* *pdfs* cross. In the usual method of constant stimuli, however, there are many fewer noise than signal stimuli. The segregation of these stimuli into pure blocks is represented by $N_{1/6}$ and signals 1, ..., 6 such that a block contains *k*/6 noise trials and *k* trials with a signal of a particular intensity *i*. The corresponding optimal criteria are the $c'_i$. (d) The probability of a detection response ("yes") as a function of signal intensity in typical mixed lists and in typical pure blocks compared: *m* indicates mixed lists (method of constant stimuli) using criterion $c_2$ as shown in (a) and (b); *p* indicates pure blocks using criteria $c_i$ shown in (c). (e) Expected probability of correct responses in mixed lists and in pure blocks. The conditions are as in (d): *m* indicates the typical mixed case of the method of constant stimuli (higher penalties for false alarms than misses, which is equivalent to equal numbers of noise and signal stimuli, $N_6$ versus $S_{1-6}$ in (a)); *p* indicates the expected outcome with the equal number of noise and signal stimuli segregated into six pure blocks.

decrement in compound tasks, relative to component tasks, even with an ideal observer (defined in the framework of SDT).

The original, theoretical rationale behind the method of constant stimuli was that the observer could maintain a constant criterion, for example, a criterion that admitted exactly 5% false alarms, independent of the values of the signal stimuli and their proportions relative to noise. In fact, this method of defining payoffs seems not to have been attempted experimentally. When payoffs are defined in terms of the values of hits, misses, false alarms, and correct rejections, the observer's criterion varies widely with the proportion of stimuli and values of outcomes in the experiment. According to SDT analysis, the method of constant stimuli yields not a unique psychometric function but one member of a family of alternative functions depending on the mixture of stimuli and on the payoffs of the experiment. The threshold (which usually is taken as the 50%

or 75% point of the psychometric function) similarly depends on payoffs and context. The threshold is a property of the idiosyncratic experimental situation in which it is measured; it is not an invariant of the sensory system.

**7.4.3. General Description of Threshold Events.** What is needed is an economical, general method of describing human responses to threshold stimuli. Signal detection theory offers a theory for two stimuli: noise and a signal of a particular intensity. For the case of equal-variance Normal probability density functions (*pdf*s), with ($\sigma^2_{S+N} = \sigma^2_N$), and independent observations, the single parameter $d'$ provides a complete description of the subject's encoded stimulus information. Any possible predictions about detection performance can be made on the basis of $d'$ alone, plus the relevant aspects of the external situation (stimulus probabilities, payoffs, etc.).

When there is more than one near-threshold stimulus, matters are much more complex. One may wish to determine the discriminability of each stimulus $s_j$ from noise, as in the preceding example (Fig. 2.26). This yields a $d'_j$ value for each stimulus. One may ask, "How does $d'$ grow with stimulus intensity?" When $d'$ increases linearly with stimulus intensity (or energy, $e$), which it must for small ranges of $e$, the slope of this line, $\partial d'/\partial e$, is a single number from which the results of many experiments can be derived. Underlying this particular formulation is a much more fundamental issue which is considered now.

Is the discriminability $d'_{jk}$ between two stimuli $s_j$ and $s_k$ given by $|(d_{jk} = d'_k - d'_j)|$, where $d'_i$ measures the discriminability of stimulus $i$ from noise and $|\;|$ represents absolute value? This formulation of the question suggests that threshold detection is a special case of discrimination or categorization, a domain that has been extensively studied. In fact, the assumptions made by Thurstone (1947) in Case V of his theory of comparative judgment are equivalent to those of SDT (i.e., stimuli spaced on a single dimension with Normal equal-variance *pdf*s). The general problem is finding a representation of stimuli in a multidimensional space where the distance between pairs of stimuli reflects their psychological distance, that is, their confusability in various experimental settings. Although SDT is concerned with microscopic portions of this space, that is, with stimuli that are very close and highly confusable, the general, cosmic scaling methods developed are applicable to the SDT microcosmos.

To show the close relation between classical psychophysics procedures and multidimensional scaling methods, consider the hypothetical auditory detection experiment described in Figure 2.26(a). The method of constant stimuli was used to generate a psychometric function from stimuli consisting of noise $S_0$ and six tones ($S_1, S_2, ..., S_6$) ordered in increasing intensity. In the classical method, the observer simply responded "detect" (1) or "nondetect" (0) after each stimulus was presented. In the more efficient modern method, the observer also gives confidence in the response; for example, 0 = definitely noise, 1 = probably noise, 2 = possibly noise, 3 = undecided, 4 = possibly signal, 5 = probably signal, 6 = definitely signal. (The confidence data are extremely useful in discriminating between theories; they can and should be collected, usually, at no extra cost.) From the seven levels of confidence it is a small leap to ask the observer to use the same seven values to identify the stimuli. This procedure (Sperling, 1965) has two great advantages: (1) the response is objectively correct or incorrect, and therefore it can be reinforced and thereby shaped, and (2) it externalizes the internal psychological dimension. That is, when the observer says "2" to a noise stimulus for which the correct (and most frequent) response is "0," it is interpreted to mean that on that particular trial, the noise produced an internal response of the same magnitude as the stimulus $S_2$ usually does (assuming that $S_2$ elicits "2" more frequently than other responses).

By a small modification in procedure, the psychophysical method of constant stimuli becomes an experiment in absolute identification in which $n$ stimuli are presented one at a time, and the subject must identify each of them. The resulting $n \times n$ (stimulus $\times$ response) confusion matrix can be analyzed by any of the many scaling methods that have been developed for this purpose. Analysis by Thurstone's Case V (1947) would be equivalent to a generalized SDT. Other such analyses are equivalent to other assumptions (than normal equal-variance *pdf*s) about encoding variability at threshold. For example, various multidimensional scaling methods derive both the *pdf* and the distance metric from stimulus–response matrices.

The application of multidimensional scaling to threshold data is especially interesting when the sensory continuum under study is not simply one dimensional. For example, brief threshold increments and decrement pulses of light are said to appear more similar to each other than to zero—no pulse at all. Sperling (quoted in Levitt, 1972, p. 160) conducted an identification experiment with near-threshold pulses to investigate this sensory continuum experimentally. His confusion data were analyzed by a multidimensional scaling program which revealed that, indeed, this sensory continuum was not a straight line. It was horseshoe shaped, with increment and decrement flashes juxtaposed at opposed ends of the shoe (Figure 2.27).

On the whole, multidimensional scaling methods are ways of representing stimuli in a space independent of the observer's strategy. Strategy affects how the encoded stimulus information is used but not how it is represented in these models. Strategy can appear as a vector of response biases or a vector of weighting factors in the models (Shepard, 1958). On the whole, the decision rules that are necessary to translate a multidimensional representation into behavioral predictions (dependent on situational factors) have not been worked out adequately. Some models, such as INDSCAL (INdividual Differences SCALing, Carroll, 1972), which have been especially adapted for dealing with between-individual differences, are also adaptable to dealing with within-individual differences on occasions where the same individual may be using different strategies. Eventually, these more general scaling methods (with very unrestrictive assumptions) may come to supplement the highly restrictive, specialized threshold theories.

**7.4.4. Pure versus Mixed Blocks in Reaction Time**

***7.4.4.1. Foreperiod Uncertainty.*** The foreperiod is the interval between a warning stimulus and the reaction stimulus in an RT task. The blocking or mixing of foreperiods has profound effects on RT (Drazin, 1961). The more accurately the subject knows exactly when a stimulus will occur, the quicker the subject can respond to it. When foreperiods are blocked (held constant) over a series of trials, the subject can prepare for the reaction stimulus at a particular moment. Catch trials are necessary in both simple and choice RT procedures to prevent anticipations that indicate the subject is timing the response (from the warning stimulus) to coincide with the end of the foreperiod interval.

There are three main determinants of the effect of foreperiods on RT.

1. When foreperiods are very short, for example, less than about 0.2 sec, the subject may have incompletely processed the

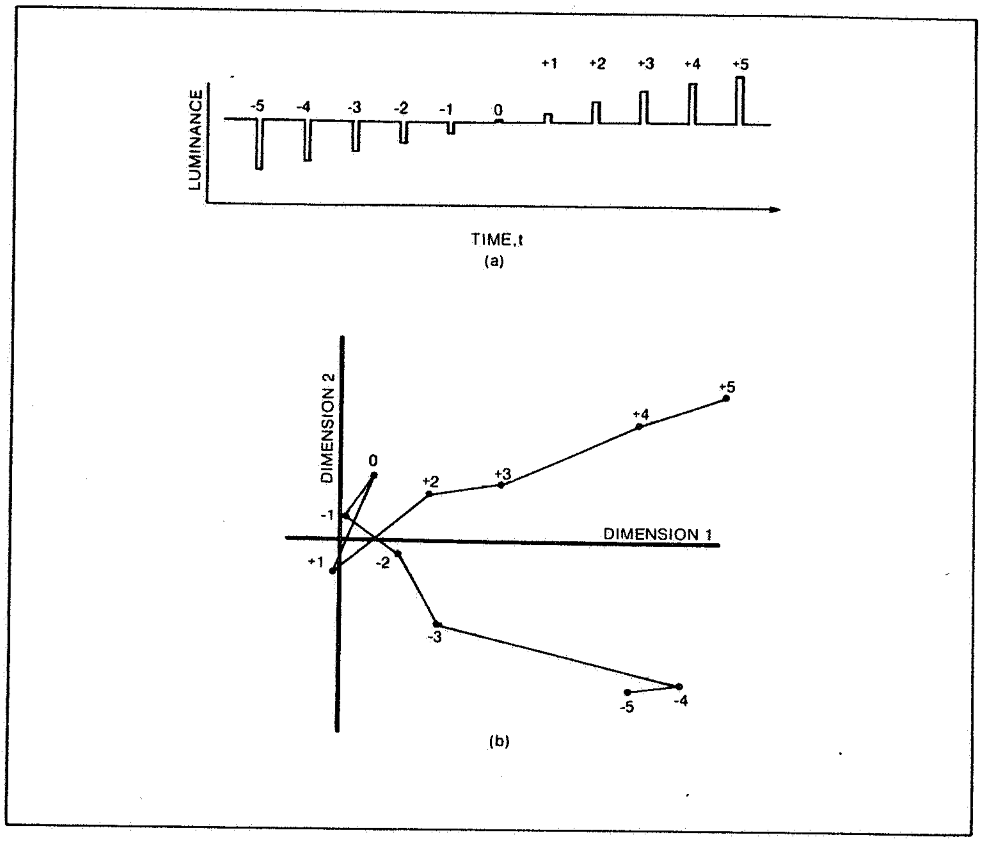



**Figure 2.27.** Scaling representation of near-threshold increments or decrements in intensity, the stimuli for a visual detection and recognition experiment. (a) The stimuli are shown schematically. The stimuli were brief pulse increments or decrements in a steady background luminance; stimulus intensity took on 11 values from −5 units to +5 units relative to the background. Subjects were required to identify which stimulus occurred on each trial. The confusion matrix resulting from this experiment was analyzed by multidimensional scaling to yield the multidimensional representation shown in (b), where close proximity in the space corresponds to greater confusion frequency. Dimension 1 represents "flashiness," independent of sign; dismension 2 represents black-to-white contrast. Large increment and decrement flashes are much more confusable with each other than can be explained by a one-dimensional (SDT) representation. (From an experiment by G. Sperling described in H. Levitt, Decision theory, signal detection theory, and psychophysics, in E. E. David and P. B. Denes (Eds.), *Human communication: A unified view*, McGraw-Hill, 1972. Reprinted with permission.)

warning stimulus when the reaction stimulus arrives; the subject will be unprepared and will produce long RTs.

2. In block designs, the longer the foreperiod in the block, the less accurately the subject can estimate the interval and, therefore, the longer the resulting RT. Under blocked foreperiods, simple RT is at a minimum when the foreperiod is near 250 msec, the time interval that optimizes the trade-off between preparedness and time uncertainty. There is only a modest increase in RT for longer foreperiods (Bevan, Hardesty, & Avant, 1965).

3. In mixed-list designs with variable foreperiods, uncertainty about when the stimulus will occur is the primary determinant of RT. This uncertainty is a function of the distribution $f(t)$ of foreperiods. The *aging*, or *hazard function*, $a(t)$ describes the probability $a(t)dt$ that the stimulus will occur in the interval $(t, t + dt)$ given that it has not yet occurred:

$$a(t) = \frac{f(t)}{1 - \int_0^t f(t')dt'} \qquad (13)$$

The warning stimulus is assumed to occur at time $t = 0$. Subjects' RT is determined, to a first approximation, by the aging function: readiness depends on expectancies.

Probably the most common distribution of foreperiods is the uniform (rectangular) distribution; its aging function increases monotonically. In the authors' experience, whenever they have used a uniform distribution of foreperiods, the RTs, conditionalized on foreperiod (the *warning* function), have shown a corresponding monotonic decrease throughout the interval. An exponential distribution of foreperiods produces a constant aging function and hence, presumably, a constant expectation of the stimulus as a function of time. In practice, however, unbounded distributions of foreperiods are impractical, so this theoretical perfection is not quite achievable. In fact, the approximately exponential foreperiod distributions that have been tested produce the fastest RTs when the mean foreperiod is about 250 msec. A review of the empirical findings is contained in Brebner and Welford (1980) and Welford (1980a).

The foreperiod effect may not be large in comparison to other stimulus and processing effects, but it is ubiquitous in

RT experiments. Although often overlooked, it should be regarded as a factor in the experimental design because partialing it out (rather than treating it as *error*) allows more sensitive statistical tests of other results. It is another example of how subjects use a priori information, about foreperiod probabilities in this instance, to optimize their performance. When such information is reduced in compound (mixed-list) designs, performance (mean RT) suffers because of the increased stimulus uncertainty.

***7.4.4.2. Intensity Effects.*** The blocking structure of experiments is also known to alter the effect of stimulus intensity on RT. When simple RT to visual stimuli of high and low intensity is measured in pure blocks, intensity may have little or no effect on mean RT (Grice, 1968; Murray, 1970). When high- and low-intensity stimuli are intermixed in the same block, there is a profound effect of intensity, with the more intense stimuli producing faster RTs. (This is but one example of the effect of experimenter-selected stimulus mixtures on RT.) These effects of intensity on RT occur with go/no-go and with variable-foreperiod procedures. The pattern of simple RT to tones of various intensities and the dependence of this pattern on the subject's knowledge have been known since the time of Wundt (1893). He presented tones of two intensities either in a strictly alternating sequence or in a random sequence within a block. These conditions produced results similar to the pure- and mixed-intensity blocks described here.

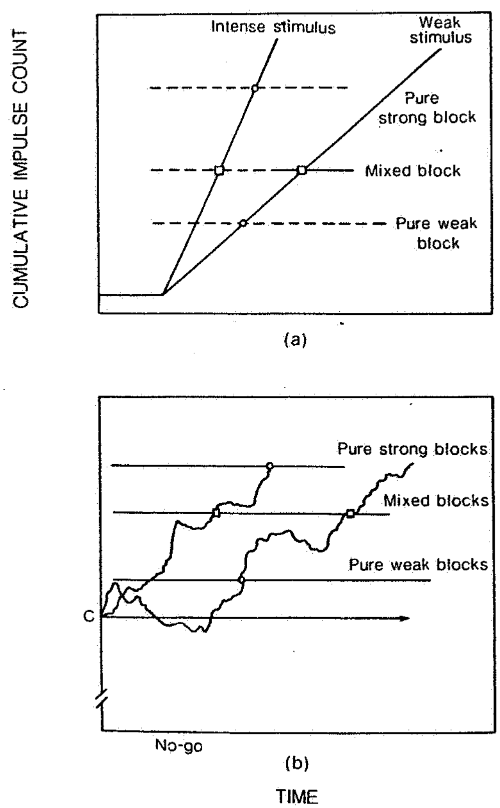



**Figure 2.28.** A schematic counting model and a random walk model equivalent to account for the effects of pure and mixed block designs on simple reaction time to weak and strong signals. (a) A schematic interpretation of the dependence of simple reaction time (RT) to weak or intense stimuli in pure versus mixed blocks. The axes are time and the cumulative count of neural impulses. Strong stimuli evoke a faster cumulation of neural impulses than weak stimuli, but detection criteria can be set to produce a large effect of intensity on RT in mixed blocks, and no effect in pure block comparisons (From Nissen (1977), Fig. 2.) (b) A random walk interpretation of the same pattern of RT data is similar to counting models that have variability in the interval between counts but differs from such counting models in that it allows negative evidence (negative counts). The axes are time and the cumulative value of the random walk. Hypothetical random walks are shown for an intense and a weak stimulus. The response boundaries are as in (a). (a. is from M. J. Nissen, Stimulus intensity and information processing, *Perception and Psychophysics*, 1977, *22*. Reprinted with permission.)

Figure 2.28(a) schematically illustrates a suggestion by Nissen (1977, after Grice, 1968) to explain RT data from Murray (1970). The subject is assumed to accumulate information from the stimulus at a fixed rate, the rate being faster for high-intensity stimuli. When a criterion amount of information is acquired, a response is initiated. The criteria in the various conditions may be set so that there is no mean RT difference between weak and intense stimuli for pure blocks but a substantial difference for mixed blocks.

The neural-counting model (McGill, 1963) and the neural-timing model (Luce & Green, 1972) are two models of sensory decision making that were designed specifically to account for intensity effects on detection and RT. The neural-timing model assumes that detection occurs with the first neural interarrival time below a criterion. In the neural-counting model, the subject is assumed to count the number of (neural) events in an observation interval, but the criterion number is assumed to be constant. Nissen's (1977) model, Figure 2.28(a), follows the spirit of the counting model but with a criterion that varies between conditions. Figure 2.28(b) schematically illustrates a random walk model which embodies similar principles but which could generate detailed, statistical predictions about RT distributions, error rates, and speed–accuracy trade-offs.

## 8. STRATEGY MIXTURES

This section examines some of the consequences of, and tests for, strategy mixtures. Mixtures between two strategies lie on a straight line in operating space. Unless the performance operating characteristic also is a straight line, strategy mixtures yield strictly worse performance than appropriately chosen pure strategies. Nevertheless, humans exhibit strategy mixture in many common tasks. Contingency tests for demonstrating strategy mixtures are described in this section. Changeover costs in switching strategies (e.g., in switching attention) lead to *path dependence* and suboptimal performance in strategy selection, a topic discussed in Section 9.

### 8.1. Definition

Assume that there exist two distinct strategies $S_a$ and $S_b$ and that the resulting performances are represented by points $a$ and $b$ on a performance operating characteristic, as shown in Figure 2.29. Suppose that a subject uses strategy $S_a$ on a fraction $\alpha$ of trials and $S_b$ on the remaining fraction $1 - \alpha$ of trials. If $\alpha$ is 0 or 1, then we say the subject is using the *pure* strategy, $S_b$ (or $S_a$). When $0 < \alpha < 1$, then the subject is using a *strategy mixture*. Whenever a strategy mixture is used, the subject's performance falls along the straight line connecting the pure strategies on the performance operating characteristic (POC). The straight-line property of strategy mixtures is obvious, because overall performance is simply an average of performances given $S_a$ and $S_b$, weighted according to the fractions of trials of each type. More generally, a mixture of a larger number of

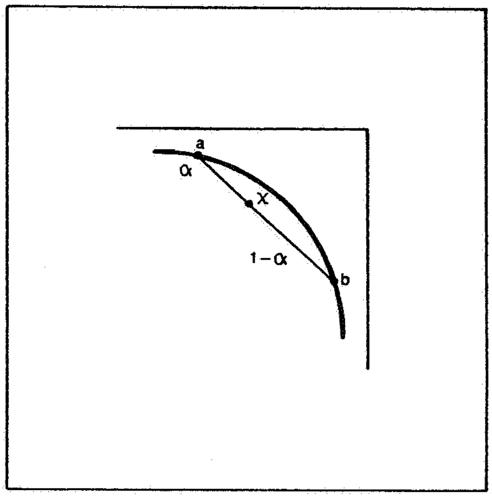



**Figure 2.29.** Operating characteristics for strategy mixtures. A performance operating characteristic with two pure strategies associated with points *a* and *b* and a strategy mixture at point *x* determined by the probability $\alpha$ of executing strategy *a* and $1 - \alpha$ of executing strategy *b*. The axes either represent performance on Task 1 versus performance on Task 2 in a concurrent experiment, or they represent two measures of performance (i.e., $P(\text{Yes}|s)$ versus $P(\text{No}/n)$ in a compound experiment. (From G. Sperling, A unified theory of attention and signal detection, in R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

strategies is represented by a point in operating space that lies at the center of gravity of the mixture. Also obvious from Figure 2.29 is that a strategy mixture results in inferior performance relative to some pure (nonpathological) strategy, except when the performance operating characteristic is a straight line, in which case the mixture yields equal performance.

Experimentally, different points along a performance operating characteristic are produced by directly instructing the subjects to change their strategy or by indirectly inducing them to change strategy in response to changes in the reward structure of the experiment. In concurrent attention experiments, the subject may be told to attend more to one or another task; in signal detection, subjects are instructed to be conservative or lax in their decisions; and in reaction-time (RT) tasks, instructions may emphasize speed or accuracy. A priori, one may not know how subjects will respond to these instructions, but one can observe the shape of the operating characteristic.

When the empirical performance operating characteristic (POC) takes the form of a straight line, it is possible that strategy mixtures are accounting for movements along the observed POC. However, this is not a necessary conclusion. Appropriate distributional assumptions can lead to a linear POC, as in the classroom example. Or the true POC may be only slightly curvilinear and hence difficult to discriminate empirically from the linear function that would result from strategy mixtures. For *concurrent tasks* there are some additional statistical tests for strategy mixtures. These are treated in the following sections.

## 8.2. Strategy Mixture in Signal Detection Experiments

**8.2.1. Two-State Threshold Models.** Threshold models are the most common source of the (implicit) assertion of strategy mixture in signal detection experiments. *Two-state threshold models* have only two sensory states, a detect state $D$ and a nondetect state $\bar{D}$. A signal detection theory (SDT) interpretation of a threshold model is illustrated in Figure 2.30. Two-state models encounter difficulties when the number of responses (e.g., confidence ratings) is larger than two. In this case, the response rule, given state $i$, is probabilistic. When the possible response alternatives are confidence ratings, a probabilistic rule means that one of several alternative responses is chosen randomly according to a probability density function (*pdf*). The set of alternative responses, and their probabilities, may be termed a *probability mix* corresponding to state $i$.

In all previous signal detection examples in this chapter, once the criterion had been chosen, there was a well-defined, optimal response for every possible observation (sensory state, $x$) on a trial. In classical SDT, a mix of response alternatives results when the criterion is chosen randomly on each trial. When the likelihood ratio $lr(x)$ is strictly monotonic in $x$, there is one and only one optimum criterion; mixing criteria means

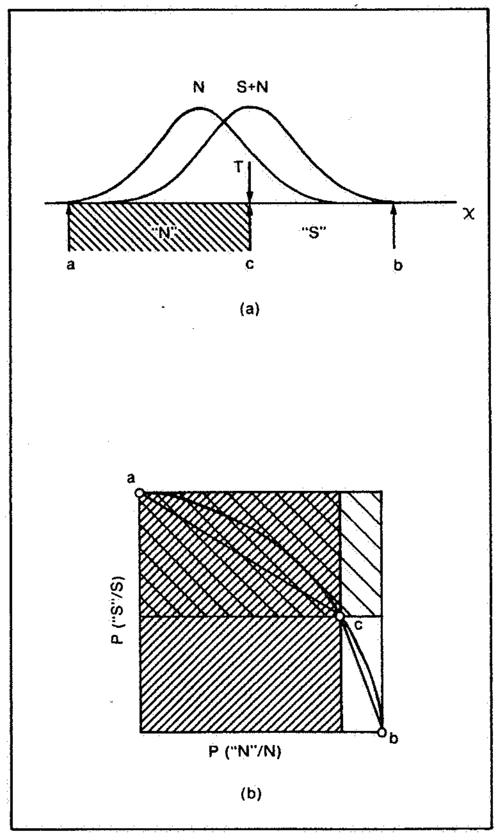



**Figure 2.30.** Graphical representations of response-threshold theory. (a) Probability density functions (*pdfs*) for noise ($N$) and for signal-plus-noise ($S + N$) stimuli on a psychological continuum ($x$). In conventional signal detection theory, the criterion ($c$) divides the continuum into "N" and "S" response regions. In response-threshold theory, the threshold ($T$) functions similarly, except that it cannot be set lower than some vlaue, resulting in a "forbidden" region, indicated by crosshatching. (b) Decision operating characteristic (DOC) for the *pdfs* in (a). Criteria *a*, *b*, *c* in panel (a) generate corresponding points *a*, *b*, *c* on the DOC in panel (b). The forbidden region for response thresholds results in a corresponding forbidden region in operating space, indicated by hatching. However, a *mixture* of strategies *a* and *c* can produce data along the straight line *ac*. In the strong form of threshold theory, the threshold can be set only at $T$ and not under or above it. The DOC is then additionally constrained to the straight line *cb*. (From G. Sperling, A unified theory of attention and signal detection, in R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

that nonoptimal criteria are used. This kind of strategy mixture inevitably degrades performance (see Figure 2.29). The reason for complicating two-state threshold models with probability mixes of responses is that, without mixes, they cannot deal with data from confidence-rating experiments. Discrete-state models produce an operating characteristic not by varying a criterion along an internal sensory continuum but by varying the probability mix of confidence ratings (or of yes-no when these are the only ratings). Under the strong assumption that the probability mix for different confidence ratings does not overlap for the two states, the two-state model produces a straight-line receiver operating characteristic with a single elbow (Krantz, 1969), and the $n$-state model could produce up to $n - 1$ elbows.

**8.2.2. $n$-State Threshold Models.** Because the two-state threshold model with nonoverlapping confidence-rating mixtures, does not give an adequate account of threshold data, two embellishments have been proposed (Krantz, 1969): (1) allow the mixes of ratings to overlap and (2) increase the number of states. Embellishment (1), the introduction of a response-confusion process, results in complex, hybrid models (with both stimulus- and response-confusion processes) that are beyond the scope of this chapter. Embellishment (2), increasing the number of allowable states, obviously can bring a discrete-state model into arbitrarily close agreement with a continuous-state model in dealing with real psychophysical data.

***8.2.2.1. Factor-Analytic Approach.*** The critical question for discrete-state models, is, What is the minimum number of states needed to account for the data? The data are entirely contained in a stimulus–response matrix that gives the probability $p(r_j|s_i)$ of a response rating $r_j$ ($j=1, \ldots, J$) given presentation of stimulus $s_i$ ($i=1, \ldots, I$). The question of the number of states reduces to a question of the rank of the $I \times J$ stimulus–response matrix, one of the most studied questions in psychometrics.

The classical approach to the rank of a matrix with imperfect data has been factor analysis (Thurstone, 1947), as developed in the context of test theory. Stimuli in the detection experiment correspond to subjects in the testing situation, and rating probabilities correspond to profiles of scores on tests, except that rating probabilities must sum to 1.0. Varying the stimulus probabilities and payoffs in the detection experiment corresponds to varying the state of the subject between test and retest in the testing situation, for example, by aging, drugs, selective motivation, differential education, and so on. The number of internal states corresponds to the number of factors needed to account for the data.

The output of a factor analysis of a rating experiment is a list of factors (probability-mix profiles), loadings (the amount of each factor in the various experimental conditions), and the percentage of variance accounted for by each factor. Whether a two-factor theory that accounts, for example, for 87% of the variance is adequate or should be supplanted by a three-factor theory that accounts for 95% of the variance is a question to which factor analysis offers no answer; it is a matter of the experimenter's and the reader's judgment. A recent statistical approach to the number-of-internal-states problem that offers significance tests (unavailable in factor analysis) has been developed by Bamber and van Santen (in press; van Santen & Bamber, 1981).

***8.2.2.2. Direct Methods.*** There are at least two alternatives to the ever-more-elaborate indirect methods of discovering whether a subject is using a complex strategy, such as randomly selecting different ratings given the same internal state. The first and always recommended procedure is to ask the subject what strategy he or she is using. A second method (Sperling, 1984) is to contrive a parallel experiment with clearly discriminable suprathreshold stimuli that induce obviously distinct internal states corresponding to the hypothesized internal states induced by the threshold experiment. Then determine whether, with these experimenter-controlled states, subjects can carry out the strategies ascribed to them with threshold stimuli. These and similar direct methods can easily be carried out concurrently with the primary signal detection study, and they may well provide useful or even critical information for the primary data analysis.

## 8.3. Strategy Mixture in Reaction Time

**8.3.1. Fast-Guess Model.** The assertion of strategy mixture in speed–accuracy trade-off comes most commonly in the guise of the fast-guess model. This model applies, for example, to two-choice RT experiments, in which the subject is presented on each trial with one of two alternative stimuli and is required to make the corresponding one of two responses as quickly as possible. For example, in Ollman's (1966) and Yellott's (1967, 1971) theory, the subject is asserted to respond to the stimulus with a stimulus-controlled RT on some fraction $(1 - \alpha)$ of the trials, and on the remaining trials $\alpha$, the subject responds as quickly as possible (simple RT) according to a predetermined guess at what the stimulus might be. On fast-guess trials, the subject is correct with only chance accuracy ($p = .5$) but with very short RTs. On the remaining trials, the subject has longer RTs and a correspondingly higher percentage of correct responses. When the experimenter demands from the subject an even lower average RT, the subject complies by increasing the proportion $\alpha$ of fast guesses.

The fast-guess model is equivalent to the assertion that the speed–accuracy trade-off (SAT) is composed of a straight-line segment whose end points represent the two strategies, the honest strategy and the fast-guess strategy. An alternative hypothesis to fast guess would be that the subject chooses a pure strategy appropriate to each payoff matrix, a process that could be modeled, for example, by boundary changes in a random walk model. This alternative strategy might generate either a curved- or a straight-line speed–accuracy trade-off. As in all the previous cases, it is not efficient to discriminate pure from mixed strategies by close examination of the curvature of the operating characteristic. In the case of the speed–accuracy trade-off, we have associated with each point on the speed–accuracy trade-off not only the mean RT and mean accuracy (which define the point) but also four RT distributions, one for each type of correct response and one for each type of error. The fast-guess model not only requires the speed–accuracy trade-off to be a straight line, but it requires the RT distribution associated with each point to be a mixture of the RT distributions associated with the extreme points. This is a powerful test to discriminate between strategy mixtures and pure strategies and is discussed in Section 8.8.1.3.

**8.3.2. Ordered Memory Scanning.** In the fast-guess model of choice RT, it is not clear why the subject would choose the fast-guess or observation strategy on any particular trial. Falmagne, Cohen, and Dwivedi (1975) deal with the question of how the strategy on the current trial is determined by local history on preceding trials. They propose a Markov model of choice RT, in which observed mean RTs are interpreted as resulting from mixtures of just two strategies (corresponding to two Markov states). Note that the more general approach of factor analysis (used in Section 8.2.2.1 to disentangle mixtures of confidence ratings) also is applicable here to disengage mixtures of RT distributions.

The subjects in Falmagne and colleagues' experiment were asked to discriminate a left-pointing from a right-pointing isosceles triangle and to respond "left" or "right," respectively. A new stimulus presentation followed the subject's response after only 100 msec, so there were more than 2 trials per sec, a situation designed to maximize sequential trial-to-trial dependencies. Reaction-time data were gathered for a large number of trials per subject to allow the examination of sequential dependencies in stimulus presentation order.

Stimulus presentation history for several trials preceding a critical trial was found to exert substantial effects on both mean RTs and on errors. For example, RT for a right stimulus was faster if it was preceded by several right stimuli. This pattern is shown in Figure 2.31, which displays (mean correct) RT arranged in a tree graph to represent the stimulus histories through trial $n - 3$.

***8.3.2.1. Two-State Markov Model.*** Falmagne and colleagues (1975) propose a two-state (two-strategy) model for their data. They assume that subjects compare the observed stimulus with two internal prototypes corresponding to the left and right stimulus ($L_p$ and $R_p$) *serially* and with *termination* in either the order ($L_p$, $R_p$) or ($R_p$, $L_p$). These two strategies are considered as two states of a Markov process, with the transitions between them being probabilistically determined by the stimulus on the previous trial. (Higher-level strategies, not considered here, might consist of the selection by the subject of the *probabilities* of switching from one state to the other.) Each node of the tree in Figure 2.31(a) is represented in the model by a probabilistic mixture of two latency distributions, where the proportions in the mixture are determined by the stimulus history. This simple two-state model was used to represent one subject's data, subject $P$, illustrated in Figure 2.31(a). For others, such as subject $F$, whose data are shown in Figure 2.31(b), Falmagne used more complex multistate models.

The data of subject P are graphed in Figure 2.31(c) in a two-dimensional operating space; that is, each point represents the left and right RTs, $RT_{Left}$ *vs.* $RT_{Right}$, for a particular stimulus history. The dashed line connects the second-order data points; that is, the points are conditional on the stimulus presented on the previous two trials. Data from subject E are shown in Figure 2.31(d). Although Falmagne and colleagues (1975) fit these subjects with different models, the data of the two subjects are quite similar in showing an operating characteristic that is concave *away* from the origin. This apparently pathological operating characteristic results from collapsing onto the two dimensions of the graph data that are actually embedded in a four-dimensional operating space consisting of RTs to the left and right stimuli for various stimulus histories and of the corresponding left and right accuracy levels. A two-dimensional strategy (choices of two operating levels, one on the left and one on the right speed–accuracy trade-off) is projected on the one-dimensional performance operating characteristic shown in Figures 2.31(c) and 2.31(d). In the middle of the performance operating characteristic, the subjects respond slowly and minimize errors. At the extremes of the performance operating characteristic when, because of the trial history, the subjects are willing to guess what the next stimulus will be, they adopt a riskier strategy that yields faster RTs but more errors when the unexpected stimulus occurs. The midpoint strategy is not inferior to a mixture of the end-point strategies; it exceeds the mixture strategy in "accuracy," a dimension not portrayed in the two-dimensional graphs of Figures 2.31(c) and 2.31(d).

Falmagne and colleagues (1975) fit a two-state mixture model to the data of subject $P$, and the solid line in Figure 2.31(c) is drawn through the second-order values from the model estimates. This is not quite a straight-line mixture here because the model selects separate error rates for left and right conditions, requiring that more errors be made on the less likely stimulus. Thus the probability mixtures from the two latency distributions need not be the same for the left and right conditions. Their two-state Markov model predicts a performance operating characteristic that is concave in the opposite direction from the data. (Falmagne and colleagues' more complex model assumes a mixture of four latency distributions and several error levels and fits the pattern of data reasonably well.)

***8.3.2.2. Dimensional Constraints.*** Falmagne and colleagues' (1975) simple two-state Markov model uses one parameter (the proportion of each state in the RT mixture) to fit all the different operating points on the performance operating characteristics that result from different sequences of prior trials. They use one parameter to describe movement along an operating characteristic that lies in a four-dimensional operating space (two accuracies, two RTs). A one-dimensional model requires the data to lie on a *curve* (not necessarily straight) in four-dimensional operating space. This one-dimensional constraint on the data was not investigated and may or may not have been satisfied. The two-dimensional projections of the data in Figures 2.31(c) and 2.31(d) suggest that the data do lie on a line in two-dimensional space but, unfortunately, not the line required by the Markov model. The random walk model (Link, 1975; Link & Heath, 1975) has two free parameters (two boundaries) and could, in principle, fit data that lie on a two-dimensional *surface* in four-dimensional operating space. There are no standard statistical methods for determining the dimensionality of data that are not describable by a linear model (such as factor analysis), but investigators should attempt to judge the dimensionality of their data by looking at various two-dimensional projections of the data before a commitment is made to a particular model or class of models.

***8.3.2.3. Ordered Memory Scanning in Memory-Retrieval Paradigms.*** Markov models have been proposed for more complex choice RT paradigms. Theios, Smith, Haviland, Trupman, and Moy (1973) use a buffer model (of items in short-term memory) to predict data of memory-search experiments based on Sternberg's (1966) paradigm. In this choice-reaction paradigm, the subject is presented with a single stimulus on each trial. The subject must make either a positive response (e.g., push a button with the right hand when the stimulus is chosen from a set of "positive" items) or a negative response (e.g., push a button with the left hand when the stimulus is chosen from the "negative" pool of items). Theios and colleagues experimentally varied the number and the presentation probabilities of the items in the positive and negative sets. Their model predicts both the memory-scanning data ($\overline{RT}$ is an increasing linear function of the number of memory items) and the effects of stimulus probability for both positive and negative stimuli ($\overline{RT}$ is faster for more probable stimuli).

Theios and colleagues' model consists of a memory buffer that contains a list of all the positive- and negative-stimulus items, each item being paired with its appropriate response.

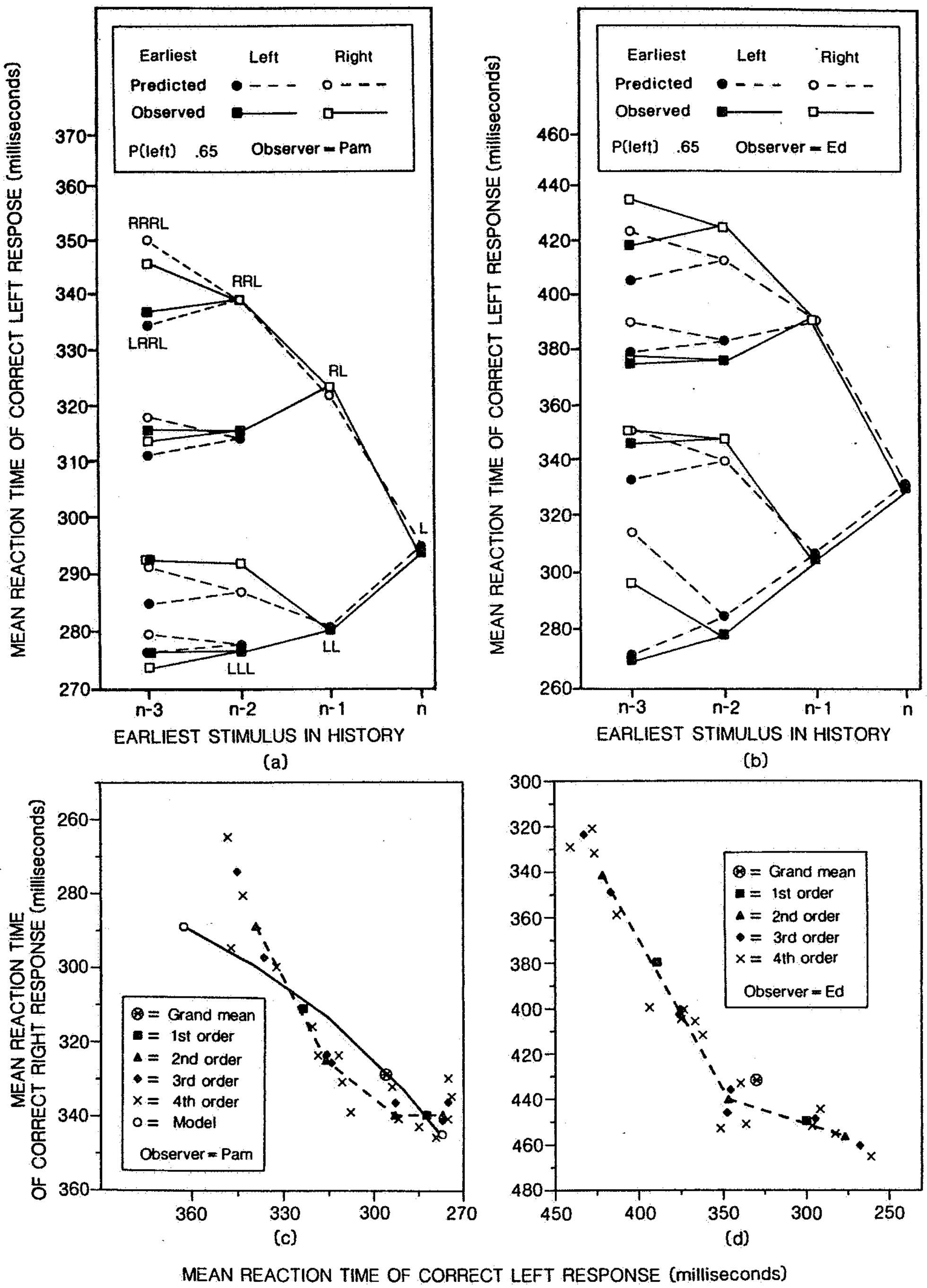



**Figure 2.31.** Tree-graph and operating-space representations of sequential dependency data from a choice reaction time (RT) experiment by Falmagne, Cohen, and Dwivedi (1975). (a) Tree graph. The abscissa is the number of the earliest preceding trial in the conditional analysis (the current trial is *n*), and the ordinate is the mean RT of a correct left response. Each node on the tree diagram is the mean correct RT conditionalized on the stimulus history of preceding trials, with third-order (three-trial) histories appearing on the left and zero order on the right. The trial history (e.g., RRRL) is indicated for seven representative nodes. Data points are filled if the earliest trial in the history is left and unfilled if it is right. Data for one stimulus, left, and one subject, Pam, are shown. (b) Same conventions and format as (a), but for subject Ed. (c) The data of panel (a), combined with the data for correct right responses, graphed in two-dimensional operating space. The axes are mean correct RT to a right stimulus and mean correct RT to a left stimulus, oriented so that fast performance is up and to the right. The dashed line connects the *empirical* second-order points which are, from top to bottom: (RRL, RRR), (LRL, LRR), (RLL, RLR), (LLL, LLR). The two open circles represent the estimated parameters of the two pure states of Falmagne and colleagues' Markov mixture model; the solid line connecting the two pure states is drawn through the four model-estimated second-order points. (d) Same format as (c), the data of subject Ed. Falmagne and colleagues used a more complex, four-state mixture model (not shown) to predict Ed's data. The "pathological" (concave-up) operating characteristics in (c) and (d) probably result from projecting a convex four-dimensional performance operating characteristic (left and right reaction times, left and right error rates) into two dimensions. (a. and b. from J. C. Falmagne, S. P. Cohen, & A. Dwivedi, Two choice reactions as an ordered memory scanning, *Attention and performance* (Vol. 5). © 1975 by Academic Press. Reprinted with permission.)

When a stimulus is presented, the buffer is scanned from top (beginning) to bottom (end) until the matching item, which contains the information defining the correct response, is found. At this point, the search terminates, and a response process is initiated. (If the matching item is not in the buffer, the subject must search a long-term memory.) This is a *serial, self-terminating* scan. Reaction time predictions are based on the length of time (about 35 msec) that it takes to compare stimulus to memory items. On each trial, there is a certain probability that the subject will change the order of items in the buffer by moving the current stimulus to the top (the most favorable scan position). In this model, each possible ordering of the buffer corresponds to a search strategy or state. (A higher-order strategy would consist of the selection of the *probability* of moving the current item to the top of the buffer and possibly the selection of the buffer length.) Overall RTs to a particular stimulus in Theios and colleagues' (1973) model are the result of a mixture of many possible states that depends on recent presentation history. This dynamic aspect of the buffer corresponds to Falmagne and colleagues' (1975) dynamic two-item buffer and is the basis of the accurate prediction of stimulus probability effects.

Raaijmakers and Shiffrin (1981) propose a model to account for recall of experimentally learned word lists. According to their model, studying a list of words results in a stored matrix of interitem strengths, where general context may also serve as an item. Upon a request to recall, the subject uses any experimenter-provided recall cues (list items and/or the context item) to prompt retrieval from memory, with retrieved items then possibly serving as retrieval cues in the next cycle of list retrieval. Recall is based on a selection of retrieved items that satisfy the recall criteria. The item or items retrieved are probabilistically determined by the stored matrix of interitem strengths and the particular cue set being used.

Raaijmakers and Shiffrin's (1981) model is even more complex than the model of Falmagne and colleagues (1975) or Theios and colleagues (1973). Performance on a given trial depends both on higher-level subject-selected strategies (i.e., number of unsuccessful iterations of the cue-retrieval cycle before stopping and the number of cue-retrieval cycles before changing the retrieval cue) and on the (probabilistic) order in which new items are recalled. Performance over trials consists of a mixture of these strategy-dependent single-trial performances.

## 8.4. Blocked versus Randomized Procedures in Speed–Accuracy Trade-off

To observe a speed–accuracy trade-off in a classical, choice RT procedure requires a blocked design in which payoffs, deadlines, or speed instructions are varied between blocks of trials. The problem with the blocked design is that strategy also varies between blocks. The experimenter cannot observe the effect of the speed manipulation on a particular strategy because the manipulation itself induces changes in strategy. For example, in a deadline experiment, the subject may use completely different strategies depending on the deadline in force. A fast deadline induces fast guesses or alternative computations, not interrupted long computations. Change of strategy is a probable confounding variable in blocked RT experiments.

The cued-response procedure (Dosher, 1976, 1979, 1981, 1982; Reed, 1973) allows different speed–acuracy trade-off conditions to be conducted in a mixed-list design. It currently is the only procedure available to obtain speed–accuracy trade-offs without the strategy confound. In this respect, it shares the good properties of the mixed-list designs (see Section 7.4) for eliminating strategy confound in partial report and in search paradigms.

However, not all strategy effects are eliminated in mixed-list designs. The memory-search and Falmagne's choice RT experiments used mixed-list designs in which all the conditions of interest are intermixed in one long sequence of trials. Nevertheless, sequential strategy effects, for example, the ordering of items in a memory buffer, were prominent. Strategy varied depending on the recent stimulus history, and these strategy variations presumably also occur in the cued-response procedure.

## 8.5. Mixed Strategies in Production and Performance

Consider the primitive plowshares-swords economy. Suppose it is decided to devote half the economy to each goal. Does it make any difference whether on every odd-numbered day of the year the whole economy is devoted to agriculture and on every even-numbered day the economy is devoted to defense production (mixed strategy) versus the case where on all days the resources are divided in half and equally devoted to each goal (pure strategy)? Certainly! The pure strategy is far more efficient in terms of the production facilities needed. But even if production facilities were not at issue, only the availability of labor, it would still be more efficient to divide labor equally on every day than to alternate days. The reason is that if even one laborer were more efficient at making plowshares than swords, it would be efficient to assign the person to the task to which he or she is better suited. By similar reasoning, with respect to any resource, a pure strategy is preferable whenever resources are not completely equal and interchangeable, with respect to the economic goals. This is the line of argument used previously to demonstrate that production possibility frontiers are always concave toward the origin. A mixed strategy does not take advantage of the curvature; it always lies closer to the origin and is of lower utility than the corresponding pure strategy. See Figure 2.29 in Section 8.1. There is inherent superiority in a pure strategy—optimal for the situation—over a mixture of less-than-optimal strategies.

Given the economic superiority of pure over mixed strategies, it is reasonable to ask, What limitation in allocation of mental processing resources prevents the utilization of pure strategies in human divided attention tasks? One possible answer is that there is a single processor or process involved in concurrent tasks and that there is a changeover delay incurred in switching this resource from task to task. Switching the resource within a trial produces unacceptable costs. An analogous problem occurs in computer time-sharing systems in switching from one user to another. There is an overhead cost (time and memory) incurred in swapping a second user's program into the central processor unit and the first user's program out into a buffer until it again gains access to the central processor unit. Trying to divide time too finely results in too many swaps per second with a corresponding, disproportionately high overhead cost. In the limit, as time is too finely divided between users, no useful work gets done, only swapping. Changeover costs have some interesting consequences in other systems as well; these are considered in Section 9.

## 8.6. Contingency Analysis of Strategies

**8.6.1. Contingency Analysis: Attendance Example.** To explore the properties of strategy mixtures in attention operating characteristics, consider an extreme classroom attendance example. Two courses are offered at precisely overlapping time periods, say noon until 1:00 P.M. See Figure 2.32(b). Suppose a student attends only Course 1 (strategy $S_1$). This student's performance is perfect on examinations for Class 1, zero for Class 2. Another student who attends only Class 2 (strategy $S_2$) has perfect performance for Class 2, zero for Class 1. To produce an equal mixture of the strategies, $S_{1+2}$, a third student ($M$) flips a fair coin each day before class to determine which class to attend. His performance with $S_{1+2}$ is 50% on examinations for each course. On the other hand, a fourth student ($P$) attends Class 1 from noon to 12:30 and Class 2 from 12:30 to 1:00, a pure strategy. Student $P$ also scores 50% on each class's examinations. How can we discriminate student $M$'s mixture of strategies from student $P$'s pure strategy?

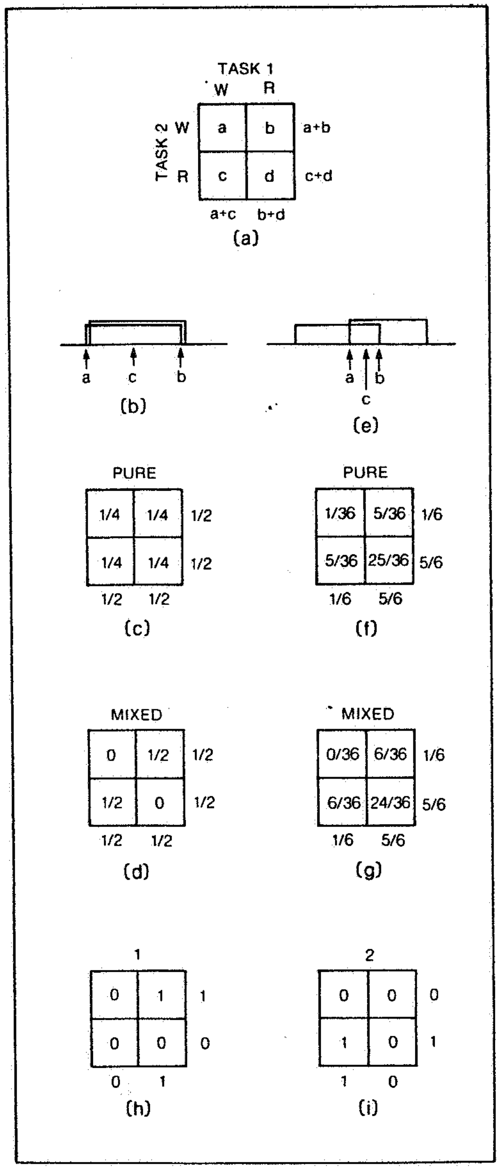



**Figure 2.32.** Contingency tests for strategy mixtures. (a) Algebraic representation of a contingency matrix for performance (wrong, right) on two tasks (1, 2); the entries ($a$, $b$, $c$, $d$) represent proportions of the total number of trials. Statistical independence of entries indicates a pure strategy; deviation from independence is evidence for switching (alternation) between strategies to attend to task 1 or task 2 on different trials (strategy mixture). (b) Information densities for a classroom example in which the two tasks (or classes) are assumed to overlap completely, yielding maximum incompatibility; $a$, $b$, $c$ represent classroom switching times within a lecture period (within a trial). (c) The contingency analysis corresponding to the situation in panel (b) with a pure strategy of changing classes at time $c$. (d) Contingency analysis corresponding to a 50 to 50% mixture over days of changing classroom at times $a$ or $b$. The matrices resulting from the two pure strategies $a$ or $b$ are shown in (h) and (i). Panels (e), (f), and (g) correspond to (b), (c), and (d), but show a situation where classes are only partly incompatible. The expected performance difference between the pure strategy (f) and strategy mixture (g) is very small (0.03) in this case.

When a performance operating characteristic (POC) is strictly concave, then a mixture of strategies lies on a straight line away from the curved POC. Insofar as an intermediate point $y$ on a POC lies above the line representing the mixture of its neighbors ($a$ and $b$), it cannot represent the mixture of strategies that gave rise to $a$ and $b$ but represents a new strategy. This test can be generalized. Suppose it is established that at least $n$ straight-line segments are required to approximate a curved POC. Then there are at least $n + 1$ different strategies. In the limit (for example, in the usual signal detection case with Normal distributions assumed for noise and signal plus noise and with a continuously variable criterion), an infinite number of possible strategies is assumed.

The problem with using the shape of the POC to infer the number, or existence, of possible intermediate strategies is that it is a statistically weak test when the POC is not very curved and it is useless when the POC is straight, as it is in the classroom example. Nevertheless, a strong differentiation between a strategy mixture and an intermediate pure strategy is possible by considering joint performance on the two tasks. For each day, consider the examination questions asked about the material covered on that day in each classroom. For simplicity, it is assumed that just one question is asked by each instructor. There are four possible outcomes of the joint response to these questions: a student can correctly answer both, neither, or one question from either one of the two classes. These outcomes are represented in the $2 \times 2$ contingency table in Figure 2.32(a). In the mixed strategy, student $M$ attends all of one class or the other. Student $M$ always answers the question from the attended class correctly, always fails the other question, never answers both questions correctly, and never misses both. See Figure 2.32(d) which illustrates performance with the mixed strategy; the performance with the component strategies of the mixture is shown in Figure 2.32(h) and (i). Over the whole examination, with questions asked about many days, student $M$'s performance will average out to 50% in each class, as shown in Figure 2.32(d).

On the other hand, if it is assumed that instructors construct their examination questions independently and that they are equally likely to probe information offered in the first half as in the second half of the class period, then student $P$, who switches classes halfway through the period, is as likely to answer any examination question as any other. Student $P$'s pure strategy results in a contingency matrix in which all cells have equal probability (see Figure 2.32(c)). Thus the pure strategy results in a contingency matrix in which there is statistical independence components, for example, for the matrices in Figure 2.32, $(d) = \frac{1}{2}(h) + \frac{1}{2}(i)$.

## 8.7. Strategy Mixture in Attention Operating Characteristics

**8.7.1. Visual Search.** The attention operating characteristics (see Figure 2.15 in Section 5.3.2) reported by Sperling and Melchner (1978a, 1978b) are nearly straight lines. The extreme strategies are "give 90% of your attention to the inside" and "to the outside," respectively. The equal-attention strategy is near the midpoint of these extremes. One may ask, Can the contingency matrix tell us whether the equal-attention strategy is a pure strategy (attention sharing) or whether it is a mixture of two extreme strategies?

The answers differ little for the three different task combinations in Sperling and Melchner. In no case were the data powerful enough to reject the switching mixture hypothesis (switching attention strategies between trials). However, the sharing hypothesis could be rejected for concurrent search for large and for small targets and for concurrent search for numerals and for letters. For the concurrent search of noise-masked and normal numerals, performance was so close to the independence point (similar to the case shown in Figures 2.32(e), 2.32(f), and 2.32(g)) that the mixed strategy and pure strategy predictions of the equal-attention matrix do not differ enough to make a discrimination feasible. Although mixture cannot be rejected for any individual subject or condition, all the data deviate somewhat from the pure mixture predictions in the direction of sharing. Thus the most likely conclusion is that strategies entering into the mixture in the equal-attention conditions are not quite as extreme as the strategies employed in "give 90% of your attention" conditions. In other words, there are more than two strategies. Equal attention is achieved by probabilistic mixing of strategies that allocate more resources to one or the other class of target, but not such an extreme allocation as is the case with the instruction: "give 90% of your attention" to one class of target.

**8.7.2. Switching Attention versus Switching Strategies.** There are two uses of the word *switch* that need to be carefully distinguished. In the classroom example, there is the switch from one class to another. This first use refers to a switch within a trial. It denotes a particular allocation of resources, time, produced by allocating time first to Class 1 and then to Class 2.

The second use of *switch* is a change between trials in the allocation of resources—a strategy switch. The contingency analysis dealt with between-trials strategy switching.

An attention switch can refer to either a within- or between-trials shift. Within a trial, an attention shift is analogous to a classroom switch. For example, if a subject always attended to the left half of an array to be searched during the first half of the processing interval and to the right half during the remainder, that would be a pure strategy involving a simple switch of attention. A between-trials switch of attention occurs, for example, when the subject sometimes attends first to the left and other times attends first to the right half of the array. Such strategy alternation represents a strategy switch (or strategy mixture).

All the psychological examples in this chapter have considered only the case of a single switch of resources, as in the classroom example. When multiple switches (e.g., back and forth between classes) are possible, the resulting performance is not *in practice* discriminable from a sharing strategy (standing in the hallway between classrooms) for the examples considered here.

## 8.8. Contingency Analysis in Speed–Accuracy Trade-off: Microtrade-offs and Conditional Accuracy Functions

Speed–accuracy performance operating characteristics result from compound (not concurrent) tasks. Nevertheless, a type of contingency analysis can be performed on speed–accuracy trade-off data, a contingency analysis examining the extent to which accuracy is related to response time *within* a given speed stress (within a single point on the operating characteristic). There are two methods for performing such an analysis: (1) conditionalizing on correct or error responses and observing mean RT (called *microtrade-off analysis* by Pachella, 1974) and (2) conditionalizing on RT and observing accuracy (called *conditional accuracy functions* by Lapin & Disch, 1972; Woods & Jennings, 1976).

### 8.8.1. Fast-Guess Model

***8.8.1.1. Microtrade-off.*** The fast-guess model (Section 8.3.1) clearly predicts that within any speed-stress condition, errors should be faster than correct responses, but observing a statistically significant difference requires a condition with reasonably high overall accuracy. This prediction is derived as follows: in the fast-guess model, all error responses are the result of fast guesses; correct responses result from a mix of stimulus-controlled processing and correct guesses. When the proportion $\alpha$ of a fast guess is very high (e.g., 1.0), correct and error reaction times (RTs) will be essentially identical since both arise from the guess RT distribution. As the proportion $\alpha$ of fast guesses declines and overall accuracy improves, the mean RT for correct and error responses should diverge. Yellott (1971) derived a more specific prediction of the fast-guess model. The weighted difference between mean correct and error RT should be linear with $p_c - p_e$.

$$p_c \overline{RT}_c - p_e \overline{RT}_e = k(P_c - p_e) \,. \tag{14}$$

Here $p$ represents probability, $\overline{RT}$ represents mean RT, and the subscripts $c$ and $e$ refer to correct and error responses, respectively. Yellott (1967, 1971) presented data from several choice RT experiments for two highly discriminable stimuli (red and green lights) which met this constraint. Whether the fast-guess model holds in any particular situation must be empirically determined.

***8.8.1.2. Conditional Accuracy Functions.*** Equation (14) describes a microtrade-off relation—mean RT as a function of accuracy. When conditional accuracy functions are measured, the fast-guess model simply predicts that accuracy should be strictly monotonically increasing with RT (Pachella, 1974).

***8.8.1.3. Fixed-Point Property.*** Although it is not a contingency analysis, a second strong test of the fast-guess model exists. Under the fast-guess model, the RT probability density function (*pdf*) for any speed-stress condition is a weighted mixture of two source *pdf*s, the fast-guess *pdf* and the stimulus-controlled *pdf*. Different speed-stresses vary only in the weighting of the two source *pdf*s. When this is true, the various speed-stress *pdf*s must observe the *fixed-point property* (Falmagne, 1968). At the point $x^*$ where the two source *pdfs* cross, the height $h^* = p(x^*)$ of the two *pdf*s is identical. Therefore, no matter what the mixture, the height of the resulting *pdf* at $x^*$ must be $h^*$. A failure of *pdf*s from different speed-stress conditions to show a fixed point $(x^*,h^*)$ constitutes a rejection of the fast-guess model. This test has generally not been applied to speed–accuracy trade-off data; it is more convenient but less powerful than the factor-analytic approach that has been recommended here for the analysis of mixtures.

***8.8.1.4. Conclusion.*** The microtrade-off prediction (faster errors than correct responses) and the conditional accuracy prediction (monotonic increasing $\overline{RT}(P_c)$) can both be tested within a single condition of speed test. For fast choice RTs to simple sensory stimuli, these predictions obtain, but they do not strongly differentiate, the fast-guess from other models. Equation (14) and the fixed-point property can be tested only between conditions. These predictions have not fared as well.

**8.8.2. Other Models.** How different are the predictions concerning microtrade-offs for other models of speed–accuracy trade-off? Although discrete-state Markov processes were introduced in Section 8.3.2, the main alternative to the fast-guess model considered in this chapter has been the random walk model. So far it has been necessary to consider random walk models only in a general way. A criterion shift (starting point or boundary) in the random walk model is a shift between pure strategies. It turns out, however, that the form of the predicted microtrade-off for a random walk model depends on the detailed assumptions of those models. The simple discrete random walk model considered by Pachella (1974) predicts equal RTs for correct responses and errors. The more complicated, discrete random walk model developed by Link and Heath (1975) assumes a possibly asymmetric distribution of steps (drift distribution) to allow faster error than correct responses. (Specifically, Link and Heath assume an asymmetric moment generating function so that the step distribution is skewed in the direction of an error.) Alternatively, Pachella noted that if a criterion of a random walk model varies from trial to trial *within* a speed stress or block (i.e., if there is strategy mixture in the random walk), this also produces faster errors than correct response. Finally, Ratcliff (1978) proposed a generalization of a continuous random walk where drift rates for "positive" and "negative" trials are normally distributed and typically overlapping. In his version of the random walk model, errors are primarily the result of the overlapping tails of the drift rate distributions, and tend to have longer RTs. Thus every possible microtrade-off pattern is consistent, at least in a general way, with some version of a random walk model. Discriminating between models requires, as usual, convergent data from many conditions and experiments.

## 8.9. Summary

The fine-grained analysis of signal detection experiments made possible an accurate and powerful description of the subject's implicit decision strategies. In the case of concurrent detection tasks, the analysis could be extended to the strategies governing the implicit allocation of mental processing resources. A similar analysis in the domain of RT experiments revealed complex, implicit decision processes that molded the subject's performance to the experimentally imposed demands. Section 9.2 demonstrates that it is possible to measure the dynamics of implicit resource allocation strategies with the same precision as one typically measures observable motor responses. Just as in physics, where elementary particles are not directly observable but can be inferred with great precision from their effects, the cognitive strategies involved in processing information are not directly observable but can be inferred and measured with mathematical precision.

# 9. PATH DEPENDENCE AND THE DYNAMICS OF STRATEGY SWITCHING

The discussion of strategies and strategy mixtures in attention operating characteristics has so far ignored the question of whether one strategy can be traded or changed for another without any changeover costs whatsoever. For example, the optimal-search models of Section 7 incorporated no changeover or sharing costs. The alternative is that a strategy switch (e.g., a switch of attention from one focus to another) entails a changeover cost. This section describes the general phenomena of path dependence that occur whenever there are changeover costs. Path dependence is widespread in studies of performance. Examples are given from partial reports in iconic memory, signal detection, and reaction time (RT) studies. These examples deal with trial-to-trial changes in strategy. Within a trial, the *attention-reaction time procedure* is shown to offer a comprehensive method for measuring the dynamics of shifting attention.

## 9.1. Path Dependence in Performance Operating Characteristics

### 9.1.1. Path Dependence in Classroom Attendance

***9.1.1.1. Path Dependence within a Single Day.*** The simplest situation in which to discover effects of changeover costs is the classroom example of Section 2. Suppose that when a student was ready to run from class 1 to class 2, the second class was located not in an adjacent room but in a different building and the trip between classes would consume 5 minutes. Clearly, there would be no point in switching from class 1 to class 2 unless the information being offered in class 2 were so much more valuable that it could compensate for the lost time.

The effect of a changeover cost is to maintain the status quo. The student remains in the present classroom, even when another class would be slightly more useful, because the additional utility is insufficient to compensate for the changeover cost. The student's current classroom reflects not only the current utility of the competing classes but also reflects the past history that brought the student to the present class. A class that was useful in the recent past holds students even after its utility has slipped below that of its competitors. The dependence of the response to the present stimulus on the response to immediately preceding stimulus is called *path dependence*.

The kind of path dependence exemplified by the persistence of previous modes of response is sometimes referred to as *hysteresis*, a reference to the electromagnetic phenomenon in which a magnetic substance tends to retain its previous magnetic orientation even after an oppositely directed external inducing field has been applied. The new field, had the previous magnetic orientation been neutral, would have been sufficient to induce a change. Of course, hysteresis can be overcome; it simply requires a stronger external inducing field. Energy is lost in a hysteresis cycle, related to the amount of path dependence, with no energy being lost when there is no hysteresis. See Figure 2.33(a).

The classroom dilemma is analogous to magnetic hysteresis. Students can be induced to switch classes, provided the required differential benefit is sufficient to overcome the cost. The classroom-switching cost, lost information during changeover, is loosely analogous to lost energy in hysteresis. (A better analogy with magnetic hysteresis is the loss in information due to the student's being in a nonoptimal classroom; see Figure 2.33(b).) When classes are adjacent and there is no changeover delay, there also is no path dependence, no hysteresis, and no lost information; see Figure 2.33(c). The student's strategy at any and every instant of time can be optimal for that instant.

There is an interesting heuristic representation of path-dependent effects in *catastrophe theory* (Thom, 1975; Thom &

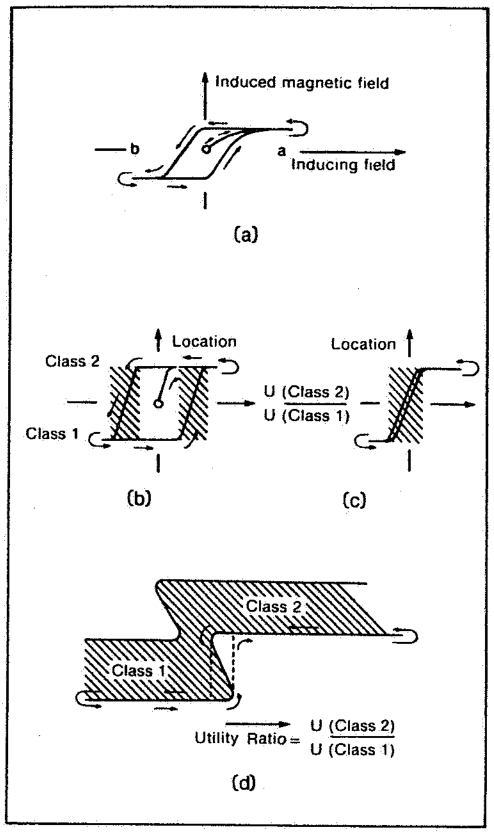



**Figure 2.33.** Schematic analysis of path dependence in classroom attendance. (a) Hysteresis in a piece of magnetized iron. The inducing field is initially neutral (open circle) and then varies back and forth between *a* and *b*, indicated on the abscissa. The ordinate indicates the induced magnetic orientation of the microcrystals in the iron. The curved arrows indicate the time sequence. (b) Hysteresis in the classroom. Two courses are offered; the ratio of their utilities, *UR* = *U*(*Class* 2)/*U*(*Class* 1) is varied periodically during a long class period. Initially (open circle) the inexperienced student is midway between classes. The information being offered in class 2 is just becoming more valuable than in class 1, so the student runs to class 2 and remains there. Subsequently, class 1 becomes more valuable, so the student runs to it. As the utility of class 1 subsides, the student returns to class 2, and so on. Information is completely lost during transit (heavily shaded area) and partially lost during the time the student lingers in the less informative class (one-half of the clear center rectangle). (c) Classroom strategy of senior student. Being smarter than an iron crystal or an entering student, the person anticipates the future course of events. When in class 2, as its utility diminishes, the student leaves while it is still more valuable than class 1, knowing that by the time he or she has arrived in class 1 the relative utility will have reversed. The person loses information only as a result of transit (shaded area), never by being in the wrong classroom. (d) Catastrophe theory representation of the events in (b). The upper surface represents class 2, the lower surface class 1. The abscissa represents the control parameter, the utility ratio, *UR* = *U*(*Class* 2)/*U*(*Class* 1). When *UR* is varied and the student reaches a fold in the surface, the student jumps to the other surface and continues there. The jump is the "catastrophe" of catastrophe theory. (From G. Sperling, A unified theory of attention and signal detection, In R. Parasuraman and R. Davies (Eds.), *Varieties of attention*, Academic Press, 1984. Reprinted with permission.)

Zeeman, 1975; Zeeman, 1976). The classroom in which we find the student may be thought of as the dependent variable, which is under the control of an independent variable, the utility ratio of the material offered in the two competing classes. As the utility ratio changes, the state changes, as described above and as illustrated in Figure 2.33(d). The catastrophe occurs when the student switches from one surface (classroom) to the other at a fold in the surface. A useful aspect of the catastrophe theory representation is that all possible equilibrium states and the relations between them are clearly shown. A limitation of the catastrophe theory representation is that neither the dynamic aspects of the situation nor the underlying processes are represented. By itself, a catastrophe theory representation is an insufficient description of a dynamic system (Sperling, 1981; Sussman & Zahler, 1978).

***9.1.1.2. Path Dependence between Days.*** The separated classrooms example illustrates how changeover costs (distance between classrooms) can cause a strategy (sitting in a particular classroom) to persist beyond the point at which it would be chosen if there were no changeover costs. A related, and more complex, phenomenon occurs in the student's choice of which classroom to enter on each day. Suppose two sessions of a course, Session 1 and Session 2, are being offered at exactly the same time. Initially, Session 1 is more informative than Session 2. With succeeding days, Session 2 improves and becomes more informative than 1. The student initially attends Session 1 and continues to attend it even after Session 2 has surpassed it. This kind of path dependence between trials has an honorable but intermittent history in the experimental psychology of attention under the pseudonyms of *set* and *Einstellung*. It, too, may be related to changeover costs—the costs to the student of discovering (by sampling the information offered in even the less advantageous session) which session currently is optimal. A related cost apparently is willingly incurred in probability matching (Restle, 1961), a phenomenon in which a subject occasionally chooses less-than-optimal gambles even when the optimal gamble could be chosen on every occasion (see Section 10).

**9.1.2. Path Dependence in Partial Report Procedures.** The example of path dependence in an iconic memory experiment was discussed in Section 7, and the data were shown in Figure 2.25. These data, which exhibit a textbook case of hysteresis, are from a subject reported in Sperling (1960). In this iconic memory experiment, cue delay was constant within a block; the cue delays were given in ascending, then descending, order for this subject, as indicated by the arrows. The subject gradually shifted from an equal-attention strategy to an attend-top-row strategy over a series of blocks but did not switch strategies soon enough. Thus there was hysteretic path dependence, a perseverance of previously appropriate strategies.

**9.1.3. Path Dependence in Signal Detection Procedures.** In an adaptive psychophysical procedure, the selection of successive stimuli is determined by the subject's responses to previous stimuli. These methods are sometimes called *up-down* procedures because the intensity of the next stimulus is varied up or down depending on whether the subject has just responded "nondetect" or "detect." The optimal rules for the determination of particular points on the psychometric function have been extensively investigated, and up-down procedures are widely used (Campbell & Lasky, 1968; Hall, 1981; Levitt, 1970; Pentland, 1980; Robbins & Monro, 1951; Taylor & Creelman, 1967).

The corresponding rules that should be adopted by a subject in setting an internal threshold criterion to optimize his or her responses have received somewhat less attention. Clearly the subject uses results of preceding trials to optimize detection parameters for the present trial (path dependence). Kac (1962, 1969) originally proposed a threshold criterion that essentially executed a random walk, adjusting appropriately up or down by a fixed amount whenever an error was made. It now appears that subjects vary their threshold criterion (nonoptimally) even when they have made a correct response and that while they move their threshold criterion in the direction of the optimum criterion (in response to changes in the probability of stimulus events), they do not move it far enough to optimize steady-state performance (Dorfman & Biderman, 1971; Kubovy & Healy, 1977; Schoeffler, 1965).

**9.1.4. Path Dependence in Reaction-Time Procedures.** In *stack* models of RT tasks, an ordered list of possible stimuli (the stack) is maintained in memory. Each stimulus item on the stack is paired with the appropriate response to be executed (see Section 8.3.2.). The order of the pairs on the stack determines the order in which comparisons of the current stimulus input to memory representations are carried out. Theios, Smith, Haviland, Trupman, and Moy (1973) propose a stack model for probability effects on RTs in memory search, and Falmagne, Cohen, and Dwivedi (1975) propose a stack model for probability effects in simple, choice RTs. In both models, the stack begins with the items in a random order and, with a probability less than 1.0, the current stimulus item causes its memory representation to advance to the head of the stack. Such a stack organizes itself to reflect a priori stimulus probabilities in terms of the mean positions of items in the stack.

As a choice RT experiment grinds inexorably on, a priori stimulus probabilities become quite well determined. The optimal strategy would appear to be to simply arrange the stack items in order of decreasing probability of occurrence and then to leave the stack unchanged until there is significant evidence that the experiment itself has been changed. Clearly this does not happen. Falmagne and colleagues' (1975) subjects change the stack according to the immediately preceding three or four stimuli, oblivious to the evidence provided by thousands of previous trials. While such a stack exhibits some hysteresis, it exhibits far too little; it is far too labile to approach optimality in a stationary environment. As in probability matching, the subject seems to give too much weight to unlikely events, although here the explanation is in terms of a decision strategy (to advance or not to advance an item in the stack) that uses no memory, being based entirely on the most recent event.

In random walk models of RT, strategy manifests itself as the choice of response thresholds for each of the alternative responses. The dynamics of how these choices are achieved have received scant attention. This is unfortunate because the dynamics of strategy choice in RT procedures, with the rich data of time, accuracy, and confidence, seem to be a promising, unexplored avenue of research.

## 9.2. Dynamics of an Attention Response

A particular state of attention is represented in a model by a particular allocation of mental processing resources; an attention shift is represented by a shift in resource allocation. The attention shift is not directly observable, but it can be inferred from its consequences. The RT of a motor response is the time from the onset of the reaction stimulus to the onset of the required response. Both the stimulus (e.g., a light flash) and the response (e.g., pressing a key) are trivial to measure. In the case of an attention response, the stimulus is easy to measure, but measuring the attention response requires ingenuity.

**9.2.1. Measuring Attention Reaction Times.** Because the attention response under study here involves "grabbing" an item from a list, the attention RT procedure is introduced by means of an analogous procedure for measuring the RT of a motor grabbing response. Figure 2.34 shows a subject seated adjacent to a conveyer belt, observing a screen upon which stimuli are displayed. The subject's task is to monitor the visual display until a visual target appears and then to reach through a small opening that gives access to the conveyer belt and to grab the first passing ball possible. The balls are numbered consecutively and arranged so that the ball numbered 1 passes the window exactly 0.1 sec after the target, the ball numbered 2 passes the window exactly 0.2 sec after the target, and so on. The subject reports the number of the ball that he or she has grabbed. From the reported number, the precise motor RT of the grabbing response can be inferred. Of course, the subject also knows the RT. To keep the subject honest, numbers on the balls must be scrambled so that, from the number, the experimenter can deduce the grabbing time but the subject can not. As a further refinement, a random one of the numbers could be omitted from all the balls on each trial. If the subject ever reported that number, it would immediately indicate a flaw in the procedure.

The subject reports the number on the ball only several seconds after it was grabbed. From the reported number, the experimenter *infers* the RT of the grabbing response, which actually occurred much earlier. The latency of the verbal report has little to do with the latency of the grabbing response; the content of the verbal report is what reveals the grabbing RT. Obviously, this is an indirect method of measuring a RT. In the case of motor RT, we have a choice of direct or indirect measures of RT. In the case of attention responses, there is no visible

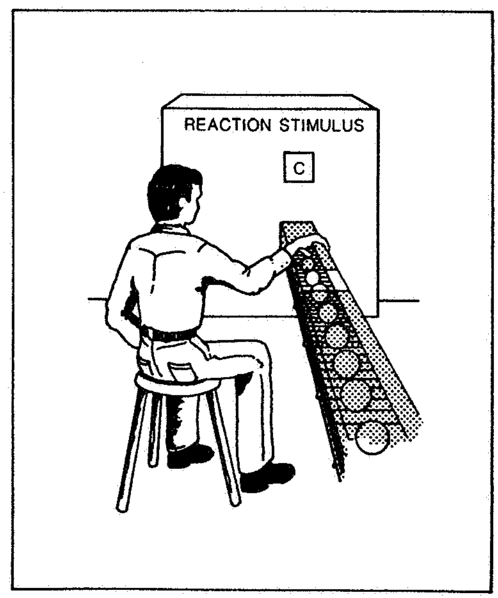



**Figure 2.34.** Measuring the reaction time (RT) of the "grabbing" response by the indirect method. The subject monitors the screen for the occurrence of the target stimulus. Instead of pressing a key as in traditional RT methods, the subject grabs the earliest possible ball from the conveyor belt. The number on the ball indicates its position in the sequence, thus providing an indirect measure of the RT. This is analogous to the indirect method used by Sperling and Reeves (1980) to measure the RT of an attention response. When a target is detected in one stimulus stream, the subject shifts attention to a stream of numeral stimuli and mentally grabs the earliest available stimulus. The subsequent verbal report of the identity of the numeral stimulus indicates its place in the sequence and, thereby, the attention RT.

response, nothing that can be directly observed; only indirect measures are possible. On the other hand, little is lost in the indirect measurement. Not only the mean but also the variance and, in fact, the whole RT distribution are obtained by the indirect method. The responses are perforce quantized into discrete times—there are balls passing only every 0.1 sec—but this is neither a serious problem nor a necessary aspect of the indirect procedure.

*9.2.1.1.* ***Mental Grabbing Response.*** To measure the reaction of a shift of visual attention, Sperling and Reeves (1976, 1978, 1980) used the following procedure. The subject maintained fixation on a fixation mark throughout a trial (see Sperling & Reeves, 1980, p. 349). To the left of fixation, a target appeared. In one series of experiments, the target was chosen at random from a letter *C*, a letter *U*, or an outline square. The target was embedded in a stream of distractors (consisting of the other letters of the alphabet) which were flashed briefly, one on top of the other, at a rate of 1 character per 110 msec. At the right of fixation, a stream of numerals occurred, one on top of the other, at either a fast rate of from 1 per 75 msec or, in other conditions, at rates as slow as 1 per 240 msec (see Figure 2.35).

The subject's task was to detect the target in the letter stream and then to report the first numeral available from the numeral stream. (In other conditions, the subjects had to report the earliest possible *four* numerals.) The task implicitly requires the subject to attend to the letter stream until the target is detected and then to shift attention to the numeral stream in order to grab the earliest numeral. The identity of the reported numeral is important only insofar as it indicates the numeral's temporal position. The attention RT on a trial is defined as the time from the onset of the target to the onset of the named numeral. From a block of trials, an entire attention RT distribution is obtained.

*9.2.1.2.* ***Critical Interval.*** There are certain important procedural considerations. In the measurement of a simple motor RT, for example, the interval between the warning stimulus, or trial initiation, and the occurrence of the target is varied randomly so that the RT experiment does not degenerate into an experiment in time estimation. Moreover, one cannot simply instruct the subject to "respond as quickly as possible." There must be explicit contingencies so that responses that occur before the target stimulus are punished as premature "anticipations"; responses occurring too late are punished for being "slow." The effect of such restrictions is to define a critical interval within which the response is supposed to occur (e.g., from 100–800 msec after target occurrence) and to reward the subject for responding as early as possible within the critical interval. Similar considerations hold in measuring the attention RT. As with the motor RT procedure, the beginning of the critical interval is placed so early that it cannot be achieved by a legitimate response. In the attention RT procedure, the numerals within the critical interval, as well as the one or two before and after it, are arranged to be all different, so that a report of a numeral's identity unambiguously indicates its position in the stream.

*9.2.1.3.* ***Simultaneous Attention and Motor Reaction Times.*** Using the motor and attention RT methods previously outlined, Reeves (1977) simultaneously measured motor and attention RTs for 3 subjects in 17 conditions, obtaining a total of over 50 pairs of attention and motor RT distributions. A representative pair of motor RT and attention RT distributions is illustrated in Figure 2.35. Although motor RT is measured by a *direct* method and attention RT by an *indirect* method, they are quite

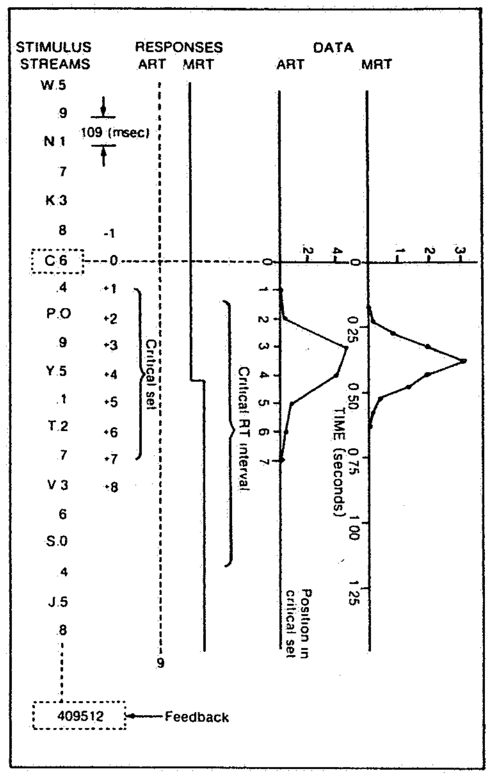



**Figure 2.35.** The results of measuring the reaction time (RT) of a shift of visual attention by the indirect method. Sample stimulus streams are shown at left. Letters in a stream are displayed one on top of the other in the same location, and the series represents time. The subject monitors the letter stream for the presence of a target, C in this case. The subject then shifts attention to the numeral stream and reports the earliest possible numeral. The seven numerals immediately following the target letter in time define the critical set, and correspond to the critical interval in a motor RT procedure. The attention RT panel shows the actual histogram of attention RTs derived from the position in the critical set of the numeral reported. The right-most panel shows a histogram of the concurrently measured motor RTs. (From G. Sperling & A. Reeves, Measuring the reaction time of an unobservable response, in R. Nickerson (Ed.), *Attention and performance VIII*, Lawrence Erlbaum, 1980. Reprinted with permission of the International Association for the Study of Attention and Performance.)

comparable in terms of their mean and variance. These data are typical of the attention RT procedure when the numeral stream occurs at a high rate (7 per sec or faster); with slow numeral rates, attention RTs become shorter than motor RTs.

Attention RTs vary in a way similar to motor RTs with manipulations of target difficulty (both motor and attention RTs get slower for hard-to-detect targets) or target probability (both motor and attention RTs get faster as the likelihood of a target increases). The indirect method yields measures of attention RTs—unobservable responses—that are no less reliable and no more difficult to obtain than directly observable motor RTs. The indirect measurement of the dynamics of a reallocation of mental processing resources depends on the *effect* of the reallocation, on how soon the reallocated resources have an effect on behavior. The indirect method can, in principle, be extended to many other attentional tasks.

**9.2.2. Order Properties in the Attention Response.** In some blocks of trials, Sperling and Reeves (1980) asked their subjects to report the first four numerals following the target, not just the earliest numerals in the critical interval. The first of the four reported numerals in the report-four procedure was equivalent in all important respects to the only reported numeral in the report-one procedure. The three remaining numerals in the response provide data that are critical for models of attentional dynamics.

There are three main properties of the responses produced under the extended report procedure: *clustering*, *disorder*, and *folding*. The clustering property refers to the fact that the positions of the four numerals reported on any trial are generally clustered around the position of the peak of the attention RT distribution described in the previous section.

To compare the order of report in the four-numeral response to the actual order of the stimulus numerals, it is necessary to define a measure of order. Let $i$ and $j$ represent the stimulus positions of two reported items. Let $iBj$ (read "i before j") represent the report of the item from position $i$ before, earlier in the response than the item from position $j$. Let $i < j$ (read "i less than j") represent the fact that $i$ occurred earlier than $j$ in the stimulus. A pair of response items is in the correct order if $i < j$ and $iBj$ or if $j < i$ and $jBi$. A comparison of stimulus and response item pairs shows that, at high numeral rates (13 per sec), almost half the response pairs are in the wrong order, and this is true for all separations of $i$ and $j$. This represents total disorder in the response. At the slowest rate (4.6 per sec), about 75% of response pairs are in the correct order.

To describe the third response property, folding, it is useful to investigate $P_{iBj}$, the probability of reporting stimulus item $i$ before item $j$, irrespective of whether $i < j$ or $j < i$. Graphs of $P_{iBj}$ for all $i$ and $j$ in the critical set are shown in Figure 2.36.

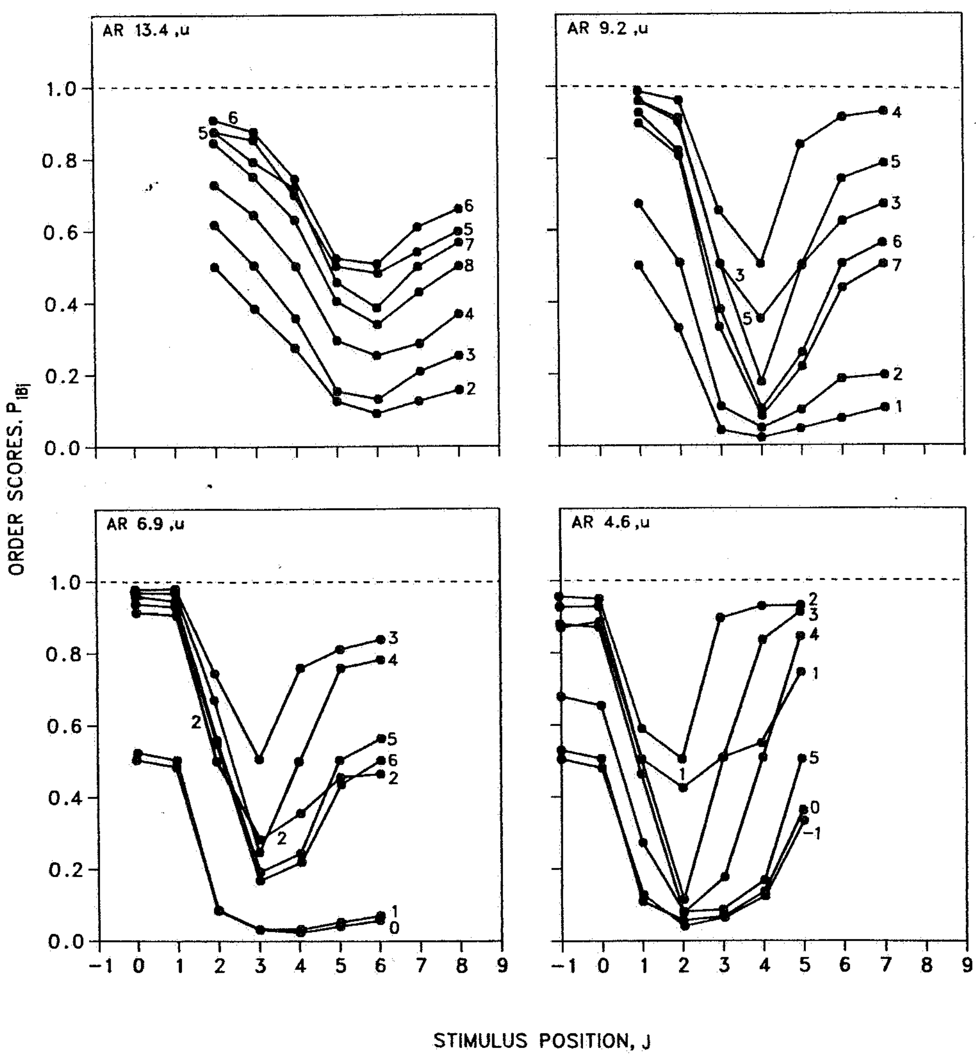



**Figure 2.36.** Order properties of the report of the first four numerals following a shift of visual attention. The experiment is identical to that outlined in Figure 2.35, except that the subject attempts to report the first four numerals following the target, rather than merely the first numeral following the target. The data shown are from one subject, AR, for four rates of display of the numeral stream (4.6, 6.9, 9.1, and 13.4 per sec) and for detection of the target *U*. The different horizontal placements of the data in various panels results from the different placements of the critical set. The ordinate, $P_{iBj}$, is the probability that the subject reports the item from stimulus position $i$ before the item from stimulus position $j$. The curve parameter is $i$, and the abscissa is $j$. Mostly, the $P_{iBj}$ data exhibit the property of laminarity, indicating that the order of report respects an internal order of strengths (e.g., 6, 5, 7, 8, 4, 3, 2 for 13.4 per sec) that is related to the objective stimulus order (2, 3, 4, 5, 6, 7, 8) by folding at position 6. (From A. Reeves & G. Sperling, Attentional theory of ordering information in short-term visual memory, *Mathematical studies in perception and cognition*, 1983, *83-7*. Reprinted with permission.)

The data exhibit the important property of *laminarity*, that is, the $P_{iBj}$ versus $j$ curves (for different values of $i$) do not cross each other.

Laminarity in Figure 2.36 is equivalent to the monotonicity property in multidimensional scaling. Provided that some other weak constraints are satisfied, monotonicity means there exists a consistent strength ordering of the stimulus positions ($i$) with the strongest position tending to be reported first, the next strongest second, and so on, and that all significant properties of the data are derivable from this ordering. The order of these strengths reveals the third property, folding. The strength (subjective order) is related to the objective order by folding. For example, in Figure 2.36, rate 9.2/sec, position 4 is strongest, followed alternately by earlier and later positions, vis-à-vis, 4, 5, 3, 6, 7, 2, 1.

A simple Thurstone Case V strength model accounts for 95% of the variance in Reeves and Sperling's (1983, 1986) data and demonstrates the usefulness of describing response order by examining pairs of items. The $P_{iBj}$ measure makes available the full power of the well-established scaling methods originally developed for pair comparisons.

**9.2.3. Gating Model of Attention Switching.** A simple *gating model of attention switching* accounts for the complex properties of clustering, disorder, and folding that occur in the report of the first four items following an attention switch. The model assumes that an attention *gate* opens after a delay time $\tau$ following target presentation. Here $\tau$ represents the time required to detect the target and to initiate an attention response; $\tau$ depends on precisely the same factors that influence motor RT: difficulty of detecting the particular target, expectancy, and so on. The attention gate opens and closes as quickly as possible, but even the fastest gate action (which follows the time course illustrated in Figure 2.37) is slow by camera standards.

Items in the numeral string that enter through the attention gate accrue strength depending on the state of the gate. An analogy can be drawn to a snapshot taken with a shutter that opens and closes gradually: the exposure of objects in the snapshot is determined jointly by how long the object is displayed and how wide open the shutter is during the display time. The snapshot is developed, and the subject reports the items from the snapshot in order of their clarity, those items with the greatest exposure being reported first, and so on. This report produces the properties of clustering, disorder, and folding. There is clustering because objects that appear close in time to the moment of maximum gate opening will tend to have maximum strength. There is disorder and folding because objects that occur early during the attention gate have low-attention weighting and are reported intermixed with late-occurring items, after the middle items.

Apart from shifting attention, strategy occurs in the gating model as a stretching out of the duration of a gate opening (beyond the minimum) to admit more items. In fact, for the

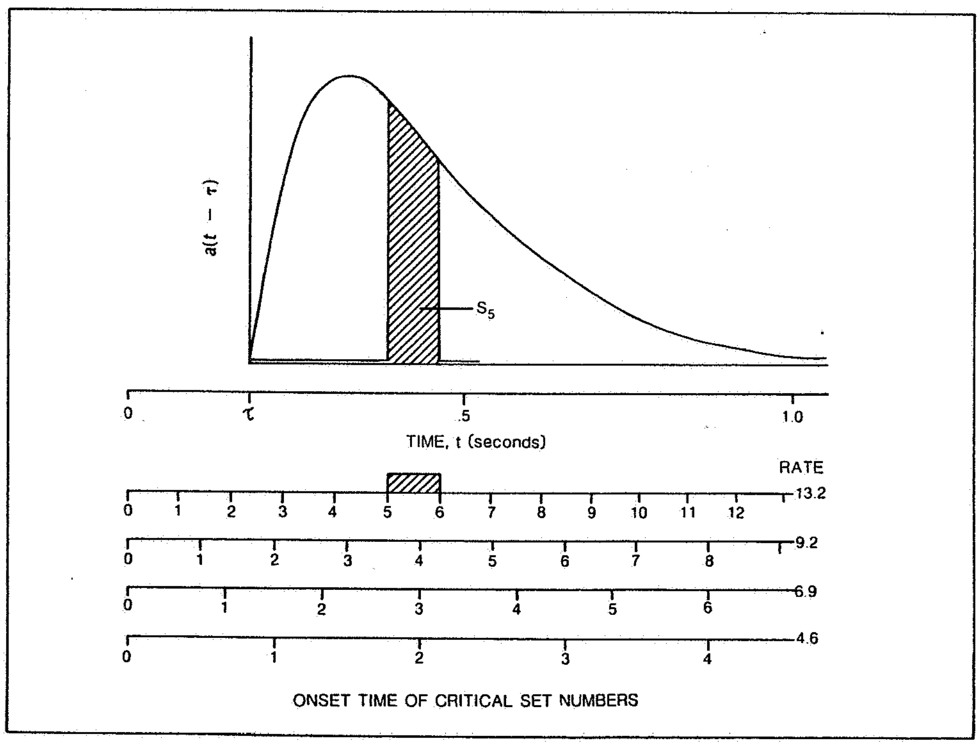



**Figure 2.37.** The gating model of the shift of visual attention from the letter (target) stream to the numeral stream for the Sperling and Reeves (1980) experiment. The bottom of the figure shows the temporal placement of the numerals at four different display rates; the period of availability of the fifth numeral at a rate of 13.2 per sec is illustrated. The model assumes that an attention gate begins to open at time $\tau$ and opens and closes gradually, as shown by the function $a(t)$. The memory strength $S_i$ of item $i$ is $S_i = \int_{t_i}^{t_{i+1}} a(t')dt'$, the total attention received from the onset of the numeral ($t_i$) to the onset of the following numeral ($t_{i+1}$), indicated for the +5 numeral by the shaded area $S_5$. Numerals are then reported in descending order of strength. This model accounts for the data of Figures 2.35 and 2.36. (From A. Reeves & G. Sperling, Attentional theory of ordering information in short-term visual memory, *Mathematical studies in perception and cognition*, 1983, *83-7*. Reprinted with permission.)

item rates (13.6–4.2 per sec) studied by Sperling and Reeves (1980) and analyzed by Reeves and Sperling (1983, 1986), this increase in gate opening did not seem to occur; stretching the moment of attention apparently requires still slower rates. The simple gating model, with only two parameters to describe the course of attention, and several to describe the detection delay $\tau$ as a function of stimulus conditions, accurately predicts hundreds of data points representing many different conditions and aspects of the data.

## 10. OPTIMIZATION?

Is all behavior arranged so as to optimize some utility function? To gain insight into the answer to this question, consider the following Gedanken experiment. A machine (or a computer program) is built so that it will optimize its behavior. Then a part of the machine is broken (or an instruction is perturbed) so that the resulting behavior becomes less than optimal. Perhaps this broken machine still could be described as optimizing a strange utility function, but whether or not this were possible, it obviously would be better described as a faulty machine that behaved nonoptimally in some circumstances.

Testing the hypothesis of optimization versus alternative hypotheses offers the same difficult problems as deciding between models in other domains, in which issues of accuracy, generality, simplicity, and efficiency must be weighed. However, the a priori evidence from considerations of biological evolution and human nature suggests that the hypothesis of optimization should be taken as a null hypothesis, the yardstick against which alternative hypotheses are matched.

For example, in two-alternative forced-choice gambling situations in which one alternative produces a reward on, say, 90% of the trials on which it is chosen and the other on, say, 30% of the trials, the optimal strategy is to always elect the highest-yield choice. Humans and other species do not do this. They occasionally elect the lower-yield choice, typically in the same 1:3 ratio as the reward ratio in this example, one of a large class of phenomena often described as probability matching (Estes, 1976; Herrnstein, 1974; Luce, 1959; Prelec, 1982; Restle, 1961; Thomas, 1975). While probability matching is a nonoptimal strategy under the very narrow constraints of the experimentally defined situation, in natural situations, as well as in encounters with real psychologist-experimenters, things do not always remain the way they seemed at first. In fact, when subjects fail to sample and to notice an alternative that the experimenter has cunningly advanced in relative value, they are convicted of "persistence of set," or *Einstellung* (Section 9.1.1.2). It is prudent to probe the environment to assess the current situation, even when such confirmatory information has a cost, as it does in probability-matching situations.

For optimum performance in choice reaction-time (RT) experiments with two alternatives, subjects should always prepare for the most likely alternative. Falmagne and colleagues (1975) found that their subjects did not do so (Section 8.3.2). The subjects' RTs depended on the stimuli that happened to have been presented in the immediately preceding three or four trials. Overall, the subjects' preparation for one or the other response alternative greatly resembles probability matching as observed in prediction experiments (Restle, 1961, chapter 6).

Falmagne and colleagues propose a two-state Markov model to account for their RT data. The probability matching in the differential response preparation results from the limited memory being used in decision making. The choice of strategy is limited to two alternative strategies (prepare for left, prepare for right) and the choice depends directly only on the last stimulus. All the information the subject may have acquired about stimulus probabilities from the observation of all the previous experimental events is represented by one of only two possible memory states. Here probability matching derives not from wisdom about the world but from meager computation.

Whether simple models (such as the two-state Markov model just described) that ignore optimization constraints in complex situations can adequately describe behavior is an unresolved empirical matter. And whether behavior, even if adequately described, can be understood without reference to optimization constraints is a matter of taste. However, intelligent behavior is seldom simple and seldom remains adequately described by a simple model once it has been intelligently investigated. Models that do not take expectations and uncertainty into account are likely to have as restricted a scope in the laboratory as they do in the real world.

## REFERENCE NOTES

1. Sperling, G., & Harris, J. R. Unpublished experiments. Bell Laboratories, Murray Hill, N.J., 1976–1977.
2. Vorberg, D. *Bayesian estimation of arbitrary points on psychometric functions.* Presented at the 13th annual Mathematical Society Meeting, Madison, Wis., 1980.

## REFERENCES

* References preceded by an asterisk are "key references."

Allport, D. A., Antonis, B., & Reynolds, P. On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, 1972, *24*, 225–235.

Anstis, S. M. A chart demonstrating variations in acuity with retinal position. *Vision Research*, 1974, *14*, 589–592.

*Audley, R. J. Some observations on theories of choice reaction time: Tutorial review. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.

Bamber, D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 1975, *12*, 387–415.

Bamber, D., & van Santen, J. P. H. How many parameters can a model have and still be testable? *Journal of Mathematical Psychology*, in press.

Bashinski, H. S., & Bacharach, V. R. Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Perceptions and Psychophysics*, 1980 *28*, 241–248.

Bevan, W., Hardesty, D. L., & Avant, L. L. Response latency with constant and variable interval schedules. *Perceptual and Motor Skills*, 1965, *20*, 969–972.

*Bouma, H. Visual search and reading: Eye movements and functional visual field: A tutorial review. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Erlbaum, 1978.

Brebner, G. M. T., & Welford, A. T. In A. T. Welford (Ed.), *Reaction times*. New York: Academic Press, 1980.

Broadbent, D. E. The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 1954, *47*, 191–196.

Broadbent, D. E. *Perception and communication*. London: Pergamon Press, 1958.

Budiansky, J. T., & Sperling, G. *GSLetters. A general purpose system for producing visual displays in real time and for running psychological experiments on the DDP24 computer.* Unpublished technical memorandum, Bell Laboratories, 1969.

Campbell, R. A., & Lasky, E. Z. Adaptive threshold procedures: BUDTIF. *Journal of the Acoustical Society of America*, 1968, *44*, 537–541.

*Carroll, J. D. Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. 1). New York: Seminar Press, 1972.

Cherry, E. C. Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 1953, *25*, 975–979.

Cherry, E. C., & Taylor, W. K. Some further experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America*. 1954, *26*, 554–559.

Cohn, T., & Lasley, D. Detectability of a luminance increment: Effect of spatial uncertainty. *Journal of the Optical Society of America*, 1974, *64*, 1715–1719.

Craik, K. J. W. Theory of the human operation in control systems: II. Man as an element in a control system. *British Journal of Psychology*, 1948, *38*, 142–148.

Davis, F. T., Kramer, P., & Graham, N. Uncertainty about spatial frequency, spatial position, or contrast of visual patterns. *Perception and Psychophysics*, 1983, *33*, 20–28.

Dorfman, D. D., & Biderman, M. A leaning model for a continuum of sensory states. *Journal of Mathematical Psychology*, 1971, *8*, 264–284.

Dosher, B. A. The retrieval of sentences from memory: A speed-accuracy study. *Cognitive Psychology*, 1976, *8*, 291–310.

Dosher, B. A. Empirical approaches to information processing: Speed accuracy tradeoff functions or reaction time. *Acta Psychologica*, 1979, *43*, 347–359.

Dosher, B. A. The effects of delay and interference: A speed-accuracy study. *Cognitive Psychology*, 1981, *13*, 551–582.

Dosher, B. A. Effect of sentence size and network distance on retrieval speed. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1982, *8*, 173–207.

Dosher, B. A. Degree of learning and retrieval speed: Study time and multiple exposures. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1984, *10*, 541–574.

Drazin, D. H. Effects of foreperiod, foreperiod variability and probability of stimulus occurrence on simple reaction time. *Journal of Experimental Psychology*, 1961, *62*, 43–50.

Due, J. F. *Intermediate economic analysis*. Chicago: Richard D. Irwin, 1951.

Duncan, J. The demonstration of capacity limitation. *Cognitive Psychology*, 1980, *12*, 75–96.

*Egan, J. P. *Signal detection theory and ROC analysis*. New York: Academic Press, 1975.

Estes, W. K. The cognitive side of probability learning. *Psychological Review*, 1976, *83*, 37–64.

Falmagne, J. C. Note on a simple property of binary mixtures. *British Journal of Mathematical and Statistical Psychology*, 1968, *21*(1), 131–132.

Falmagne, J. C., Cohen, S. P., & Dwivedi, A. Two-choice reactions as an ordered memory scanning process. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V*. London: Academic Press, 1975.

Fisher, D. L. Limited-channel models of automatic detection: Capacity and scanning in visual search. *Psychological Review*, 1982, *89*, 662–692.

Fitts, P. M. Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 1966, *71*, 849–857.

*Galambos, J. *The asymptotic theory of extreme order statistics*. New York: Wiley, 1978.

Gilliom, J. D., & Sorkin, R. D. Sequential versus simultaneous two channel signal detection: More evidence for a high level interrupt theory. *Journal of the Acoustical Society of America*, 1974, *56*, 157–164.

Glucksberg, S., & Cowen, G. N. Memory for nonattended auditory material. *Cognitive Psychology*, 1970, *1*, 149–156.

Graham, N., Kramer, P., & Yager, D. Explaining uncertainty effects and probability summation. *Investigative Ophthalmology and Visual Science*, ARVO Supplement, 1983, *24*, 186.

*Green, D. M., & Luce, R. D. Speed-accuracy tradeoff in auditory detection. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.

*Green, D. M., & Swets, J. A. *Signal detectability and psychophysics*. New York: Wiley, 1966.

Grice, G. R. Stimulus intensity and response evocation. *Psychological Review*, 1968, *75*, 359–373.

*Gumbel, E. J. *Statistics of extremes*. New York: Columbia University Press, 1958.

Hall, J. L. Hybrid adaptive procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America*, 1981, *69*, 1763–1769.

Harris, J. R., Shaw, M. L., & Bates, M. Visual search in multicharacter arrays with and without gaps. *Perception and Psychophysics*, 1979, *26*(1), 69–84.

Herrnstein, R. J. Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior*, 1974, *21*, 159–164.

Hicks, J. R., & Allen, R. G. D. A reconsideration of the theory of value. *Economica*, 1934, *1*, 52–76, 196–219.

Hirsch, J., Hylton, R., & Graham, N. Simultaneous recognition of two spatial-frequency components. *Visison Research*, 1982, *22*, 365–375.

Hoffman, J. E., & Nelson, B. Spatial selectivity in visual search. *Perception and Psycophysics*, 1981, *30*, 283–290.

Howarth, C. I., & Lowe, G. Statistical detection theory of Piper's Law. *Nature*, 1966, *212*, 324–326.

Kac, M. A note on learning signal detection. *IRE Transactions on Information Theory*, 1962, *IT-8*, 126–128.

Kac, M. Some mathematical models in science. *Science*, 1969, *166*, 695–699.

Kadane, J. B. Optimal whereabouts search. *Operations Research*, 1971, *19*, 894–904.

*Kahneman, D. *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Kahneman, D., & Treisman, A. Selection, filtering and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention*. New York: Academic Press, 1984.

*Kantowitz, B. H. Double stimulation. In B. H. Kantowitz (Ed.), *Human information processing: Tutorial in performance and cognition*. Hillsdale, N.J.: Erlbaum, 1974.

Karlin, I., & Kestenbaum, R. Effects of number of alternatives on the psychological refractory period. *Quarterly Journal of Experimental Psychology*, 1968, *20*, 167–178.

Kendall, M. G. *Rank correlation methods* (4th ed.). London: Griffin, 1970.

Kinchla, R. A. The role of structural redundancy in the detection of visual targets. *Perception and Psychophysics*, 1977, *22*, 19–30.

Kinchla, R. A. The measurement of attention. In R. S. Nickerson (Ed.), *Attention and performance VIII*. Hillsdale, N.J.: Erlbaum, 1980.

Kinchla, R. A., & Collyer, C. E. Detecting a target letter in briefly presented arrays: A confidence rating analysis in terms of a weighted additive effects model. *Perception and Psychophysics*, 1974, *16*, 117–122.

Koopman, B. O. The theory of search: I. Kinematic bases. *Operations Research*, 1956, *4*, 324–346. (a)

Koopman, B. O. The theory of search: II. Target detection. *Operations Research*, 1956, *4*, 503–531. (b)

*Koopman, B. O. The theory of search: III. The optimum distribution of searching effort. *Operations Research*, 1957, *5*, 613–626.

*Krantz, D. H. Threshold theories of signal detection. *Psychological Review*, 1969, *76*, 308–324.

Krueger, L. E. A theory of perceptual matching. *Psychological Review*, 1978, *85*, 278–304.

Kruskal, J. B. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society*, Series B, 1965, *27*, 251–263.

Kubovy, M., & Healy, A. F. The decision rule in probabilistic categorization: What it is and how it is learned. *Journal of Experimental Psychology: General*, 1977, *106*, 427–446.

LaBerge, D. Acquisition of automatic processing in perceptual and associative learning. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V*. London: Academic Press, 1975.

Laming, D. R. *Information theory of choice-reaction times*. New York: Academic Press, 1968.

Lapin, J. S., & Disch, K. The latency operating characteristic: I. Effect of stimulus probability on choice reaction time. *Journal of Experimental Psychology*, 1972, *92*, 116–427.

Levine, M. V. Transformations that render curves parallel. *Journal of Mathematical Psychology*, 1971, *7*, 410–444.

Levitt, H. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 1970, *49*, 467–477.

Levitt, H. Decision theory, signal detection theory, and psychophysics. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill, 1972.

Link, S. W. The relative judgment theory of two-choice response time. *Journal of Mathematical Psychology*, 1975, *12*, 114–135.

*Link, S. W. New views of reaction time and accuracy. In N. J. Castellan, Jr., & F. Restle (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, N.J.: Erlbaum, 1978.

Link, S. W., & Heath, R. A. A sequential theory of psychological discrimination. *Psychometrika*, 1975, *40*, 77–105.

Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.

Luce, R. D., & Green, D. M. A neural timing theory for response times and the psychophysics of intensity. *Psychological Review*, 1972, *79*, 14–57.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. The psychophysics of categorical perception. *Psychological Review*, 1977, *84*, 452–471.

McGill, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.

Mertens, J. J. Influence of knowledge of target location upon the probability of observation of peripherally observable test flashes. *Journal of the Optical Society of America*, 1956, *46*, 1069–1070.

Metz, C. E., Starr, S. J., Lusted, L. B., & Rossmann, K. Progress in evaluation of human observer visual detection performance using the ROC curve approach. In C. Raynaud & A. Todd-Pokropek (Eds.), *Information processing in scintigraphy*, Orsay, France: Commissariat a l'Energie Atomique, Departement de Biologie, Service Hospitalier Frederic Joliot, 1975.

Moray, N. Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 1959, *11*, 56–60.

Moray, N. *Listening and attention*. London: Penguin, 1969.

Mowbray, G. H. Perception and retention of verbal information presented during auditory shadowing. *Journal of the Acoustical Society of America*, 1964, *36*, 1459–1465.

Murphy, B. J. Pattern thresholds for moving and stationary gratings during smooth eye movements. *Vision Research*, 1978, *18*, 521–530.

Murphy, B. J., Kowler, E., & Steinman, R. M. Slow oculomotor control in the presence of moving backgrounds. *Vision Research*, 1975, *15*, 1263–1268.

Murray, H. G. Stimulus intensity and reaction time: Evaluation of a decision theory model. *Journal of Experimental Psychology*, 1970, *84*, 383–391.

*Navon, D., & Gopher, D. On the economy of the human-processing system. *Psychological Review*, 1979, *86*, 214–255.

Neisser, U. Decision time without reaction time: Experiments in visual scanning. *American Journal of Psychology*, 1963, *76*, 376–385.

Neisser, U. Visual search. *Scientific American*, 1964, *210*, 94–102.

*Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1966.

Neisser, U., Novick, R., & Lazar, R. Searching for ten targets simultaneously. *Perceptual and Motor Skills*, 1963, *17*, 955–961.

Nissen, M. J. Stimulus intensity and information processing. *Perception and Psychophysics*, 1977, *22*, 338–352.

Nolte, L. W., & Jaarsma, D. More on the detection of one of *m* orthogonal signals. *Journal of the Acoustical Society of America*, 1967, *41*, 497–505.

Noreen, D. L. Optimal decision rules for some common psychophysical paradigms. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology*. Providence, R.I.: Society of Industrial and Applied Mathematics—American Mathematical Society (SIAM-AM) Proceedings, *13*, 1981.

Norman, D. A. Toward a theory of memory and attention. *Psychological Review*, 1968, *75*, 522–36.

*Norman, D. A., & Bobrow, D. G. On data-limited and resource-limited processes. *Cognitive Psychology*, 1975, *7*, 44–64.

Ollman, R. Fast guesses in choice reaction time. *Psychonomic Science*, 1966, *6*, 155–156.

*Pachella, R. G. The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Potomac, Md.: Erlbaum, 1974.

Pachella, R. G., & Fisher, D. Hick's Law and the speed-accuracy tradeoff in absolute judgment. *Journal of Experimental Psychology*, 1972, *92*, 378–384.

Pareto, V. *Manuel d'economie politique* (A. Bonnet, trans.). Paris: Giard & Briere, 1909.

Pastore, R. E., & Sorkin, R. D. Simultaneous two-channel signal detection: I. Simple binaural stimuli. *Journal of the Acoustical Society of America*, 1972, *51*, 544–551.

Pentland, A. Maximum likelihood estimation: The best PEST. *Perception and Psychophysics*, 1980, *28*, 377–379.

Pohlmann, L. D., & Sorkin, R. D. Simultaneous three-channel signal detection: Performance and criterion as a function of order of report. *Perception and Psychophysics*, 1976, *20*, 179–186.

Posner, M. I., Nissen, M. J., & Ogden, W. C. Attended and unattended processing modes: The role of set for spatial location. In H. I. Pick, Jr., & E. Saltzman (Eds.), *Modes of perceiving and processing information*. Hillsdale, N.J.: Erlbaum, 1978.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. Attention and the detection of signals. *Journal of Experimental Psychology: General*, 1980, *109*, 160–174.

Prelec, D. Matching, maximizing, and the hyperbolic reinforcement feedback function. *Psychological Review*, 1982, *89*, 189–230.

Raaijmakers, J. G. W., & Shiffrin, R. M. Search of associative memory. *Psychological Review*, 1981, *88*, 93–134.

Rachlin, H., Battalio, R., Kagel, J., & Green, L. Maximization theory in behavioral psychology. *The Behavioral and Brain Sciences*, 1981, *4*, 371–417.

*Rachlin, H., & Burkhard, B. The temporal triangle: Response substitution in instrumental conditioning. *Psychological Review*, 1978, *85*, 22–47.

Rachlin, H., Green, L., Kagel, J. H., & Battalio, R. C. Economic demand theory and psychological studies of choice. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10). New York: Academic Press, 1976.

Rachlin, H., Kagel, J. H., & Battalio, R. C. Substitutability in time allocation. *Psychological Review*, 1980, *87*, 355–374.

*Ratcliff, R. A theory of memory retrieval. *Psychological Review*, 1978, *85*, 59–108.

Ratcliff, R. A note on modeling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology*, 1980, *21*, 178–184.

Reed, A. V. Speed-accuracy trade-off in recognition memory. *Science*, 1973, *181*, 574–576.

Reeves, A. *The detection and recall of rapidly displayed letters and digits*. Unpublished doctoral dissertation, City University of New York, 1977.

Reeves, A., & Sperling, G. Attentional theory of order information in short-term visual memory. *Mathematical studies in perception and cognition* (83-7). New York University, Department of Psychology, 1983.

Reeves, A., & Sperling, G. Attentional gating in short-term visual memory. *Psychological Review*, 1986, *92*, in press.

Restle, F. *Psychology of judgment and choice: A theoretical essay*. New York: Wiley, 1961.

Robbins, H., & Monro, S. A stochastic approximation method. *Annals of Mathematical Statistics*, 1951, *22*, 400–407.

Rubenstein, H., Garfield, L., & Millikan, J. A. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 487–494.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1971 *10*, 645–657. (a)

Rubenstein, H., Lewis, S. S., & Rubinstein, M. A. Homographic entries in the internal lexicon: Effects of systematicity and relative frequency of meanings. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 57–62. (b)

Rumelhart, D. E. A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 1970, *7*, 191–216.

*Samuelson, P. A. *Economics* (11th ed.). New York: McGraw-Hill, 1980.

*Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 1977, *1*, 1–66.

Schoeffler, M. S. Theory for psychophysical learning. *Journal of the Acoustical Society of America*, 1965, *37*, 1124–1133.

Schuckman, H. Attention and visual threshold. *American Journal of Optometry and Archives of the American Academy of Optometry*, 1963, *40*, 284–291.

Shaw, M. L. A capacity allocation model for reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 586–598.

Shaw, M. L. Identifying attentional and decision making components in information processing. In R. S. Nickerson (Ed.), *Attention and performance VIII*. Hillsdale, N.J.: Erlbaum, 1980.

*Shaw, M. L. Attending to multiple sources of information: (1) The integration of information in decision making. *Cognitive Psychology*, 1982, *4*, 353–409.

Shaw, M. L., Mulligan, R. M., & Stone, L. D. Two-state versus continuous-state stimulus representations: A test based on attentional constraints. *Perception and Psychophysics*, 1983, *33*(4), 338–354.

Shaw, M. L., & Shaw, P. Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 201–211.

Shaw, P. Processing of tachistoscopic displays with controlled order of characters and spaces. *Perception and Psychophysics*, 1969, *6*, 257–266.

Shepard, R. N. Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, 1958, *55*, 509–523.

*Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 1977, *84*, 127–189.

*Smith, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, *69*, 77–110.

Sorkin, R. D., & Pastore, R. E. Stimulus binaural signal detection: Comments on time sharing in auditory perception. *Journal of the Acoustical Society of America*, 1971, *49*, 1319.

Sorkin, R. D., Pastore, R. E., & Pohlman, L. D. Simultaneous two-channel signal detection: II. Correlated and uncorrelated signals. *Journal of the Acoustical Society of America*, 1972, *51*, 1960–1965.

Sorkin, R. D., Pohlmann, L. D., & Gilliom, J. D. Simultaneous two-channel signal detection: III. 630- and 1400-Hz signals. *Journal of the Acoustical Society of America*, 1973, *53*, 1045–1050.

Sorkin, R. D., Pohlmann, L. D., & Woods, D. D. Decision interaction between auditory channels. *Perception and Psychophysics*, 1976, *19*, 290–295.

*Sperling, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, *74*, (11, Whole No. 498).

Sperling, G. A model for visual memory tasks. *Human Factors*, 1963, *5*, 19–31.

*Sperling, G. Temporal and spatial visual masking: I. Masking by impulse flashes. *Journal of the Optical Society of America*, 1965, *55*, 541–559.

Sperling, G. Successive approximations to a model for short-term memory. *Acta Psychologica*, 1967, *27*, 285–292.

Sperling, G. The search for the highest rate of search. *Symposium on Attention and Performance V.*, Stockholm, Sweden: Saltsjobaden, August 1973.

Sperling, G. Multiple detections in a brief visual stimulus: The sharing and switching of attention. *Bulletin of the Psychonomic Society*, 1975, *9*, 427.

Sperling, G. Mathematical models of binocular vision. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology*. Providence, R.I.: Society of Industrial and Applied Mathematics—American Mathematical Society (SIAM-AM) Proceedings, *13*, 1981.

Sperling, G. Unified theory of attention and signal detection. *Mathematical studies in perception and cognition* (83-3). New York University, Department of Psychology, 1983.

*Sperling, G. A unified theory of attention and signal detection. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention*. New York: Academic Press, 1984.

Sperling, G., Budiansky, J., Spivak, J. G., & Johnson, M. C. Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. *Science*, 1971, *174*, 307–311.

Sperling, G., & Melchner, M. J. Estimating item and order information. *Journal of Mathematical Psychology*, 1976, *13*, 192–213. (a)

*Sperling, G., & Melchner, M. J. Visual search and visual attention. In V. D. Glezer (Ed.), *Information processing in visual system*. Proceedings of the Fourth Symposium of Sensory System Physiology. Leningrad, U.S.S.R.: Academy of Sciences, Pavlov Institute of Physiology, 1976. (b)

Sperling, G., & Melchner, M. J. The attention operating characteristic: Some examples from visual search. *Science*, 1978, *202*, 315–318. (a)

Sperling, G., & Melchner, M. J. Visual search, visual attention, and the attention operating characteristic. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Erlbaum, 1978. (b)

Sperling, G., & Reeves, A. Reaction time of an unobservable response. *Bulletin of the Psychonomic Society*, 1976, *10*, 247.

Sperling, G., & Reeves, A. Measuring the reaction time of a shift of visual attention. *Investigative Ophthalmology and Visual Science*, ARVO Supplement, 1978, *17*, 289.

Sperling, G., & Reeves, A. Measuring the reaction time of an unobservable response: A shift of visual attention. In R. Nickerson (Ed.), *Attention and performance VIII*. Hillsdale, N.J.: Erlbaum, 1980.

Sperling, G., & Reeves, A. Gating model of visual attention. *Bulletin of the Psychonomic Society*, 1983, *17*, 354.

Sternberg, S. High speed scanning in human memory. *Science*, 1966, *153*, 652–654.

*Stone, L. D. *Theory of optimal search*. New York: Academic Press, 1975.

Stroop, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, *18*, 643–662.

Sussman, H. J., & Zahler, R. S. Catastrophe theory as applied to the social and biological sciences: A critique. *Synthese*, 1978, *37*, 117–216.

*Swets, J. A. (Ed.). *Signal detection and recognition by human observers: Contemporary readings*. New York: Wiley, 1964.

Swets, J. A. The relative operating characteristic in psychology. *Science*, 1973, *182*, 990–1000.

*Swets, J. A. Mathematical models of attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention*. New York: Academic Press, 1984.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. Decision processes in perception. *Psychological Review*, 1961, *68*, 301–340.

Taylor, M. M., & Creelman, C. D. PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 1967, *41*, 782–787.

Theios, J., Smith, P., Haviland, S., Trupman, J., & Moy, M. Memory scanning as a serial self-terminating process. *Journal of Experimental Psychology*, 1973, *97*, 323–336.

Thom, R. *Structural stability and morphogenesis* (D. H. Fowler, trans.). Reading, Mass.: W. A. Benjamin, 1975.

Thom, R., & Zeeman, E. C. Catastrophe theory: Its present state and future perspectives. In A. Manning (Ed)., *Proceedings of a Symposium Held at the University of Warwick 1973/74*. In Dynamical Systems—Warick 1974. Berlin: Springer-Verlag, 1975.

Thomas, E. A. C. Criterion adjustment and probability matching. *Perception and Psychophysics*, 1975, *18*, 158–162.

Thurstone, L. L. *Multiple-factor analysis: A development and expansion of the vectors of mind*. Chicago: Chicago University Press, 1947.

Tognetti, K. P. An optimal strategy for a whereabouts search. *Operations Research*, 1968, *16*, 209–211.

Treisman, A. M. Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*. 1964, *3*, 449–459.

Treisman, A. M. Strategies and models of selective attention. *Psychological Review*, 1969, *76*, 282–299.

van Santen, J. P. H., & Bamber, D. Finite and infinite state confusion models. *Journal of Mathematical Psychology*, 1981, *24*, 101–111.

Vickers, D. *Decision processes in visual perception*. New York: Academic Press, 1979.

Wald, A. *Statistical decision functions*. New York: Wiley, 1950.

Wandell, B. A. Speed-accuracy tradeoff in visual detection: Applications of neural counting and timing. *Vision Research*, 1977, *17*, 217–225.

Welford, A. T. (Ed.), *Reaction times*. London: Academic Press, 1980. (a)

Welford, A. T. The single channel hypothesis. In A. T. Welford (Ed.), *Reaction times*. New York: Academic Press, 1980. (b)

Wickelgren, W. A., Corbett, A. T., & Dosher, B. A. Priming and retrieval from short-term memory: A speed accuracy analysis. *Journal of Verbal Learning and Verbal Behavior*, 1980, *19*, 387–404.

Woods, C., & Jennings, J. Speed accuracy tradeoff functions in choice reaction time: Experimental designs and computational procedures. *Perception and Psychophysics*, 1976, *19*, 92–101.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology* (rev. ed.). New York: Holt, 1954.

Wundt, W. *Grundzuge der physilogischen psychologie*. Leipzig: Engelmann, 1893.

Wundt, W. *An introduction to psychology* (R. Pintner, trans.). London: Geroge Allen & Unwin, 1912 (reprinted 1924).

Yellott, J. I. Correction for guessing in choice reaction time. *Psychonomic Science*, 1967, *8*, 321–322.

Yellott, J. I. Correction for fast guessing and the speed-accuracy tradeoff. *Journal of Mathematical Psychology*, 1971, *8*, 159–199.

Zeeman, E. C. Catastrophe theory, *Scientific American*, 1976, *234*, 65–83.