This self-satisfied bunny has just performed the operation of *self-termination*. He has been searching for a true carrot among a field occupied by many ill-tasting Carrot Creatures. Upon discovering the true carrot, our bunny is able to cease his search (self-termination) rather than *exhaustively* probing the entire field.

7 *Self-terminating vs. exhaustive search strategies*

Consider the situation where an observer is searching, either internally or externally, through some stimulus pattern for some specific subset of stimuli. Such phenomena are very easily imagined and are widely studied by psychologists. They might possibly occur in reading behavior, pattern recognition experiments, short-term memory studies, auditory discrimination training, and a host of others. Hence, it is obvious that careful investigation of this search process can possibly benefit many areas of psychology.

The task is the familiar one of deducing the contents of the intervening black box from the input and output vectors. The processing system can be imagined to operate in many different ways. For instance, the process is said to be self-terminating if the processing mechanism is capable of halting the search as soon as the critical subset is discovered (i.e., processed). This implies, of course, that on the majority of trials processing will be terminated (i.e., the critical subset discovered) before all of the stimulus pattern is processed. If the system is incapable of halting the search and must always process the entire stimulus pattern, then processing is said to be exhaustive. Self-terminating models have achieved some intuitive appeal since they appear more efficient with regards to energy conservation. Testing paradigms are scarce, and to date no foolproof method exists for discriminating between the two strategies. In addition, of the techniques that do exist, there appears to be some confusion as to their validity, or more particularly as to the strength of the assumptions they postulate.

Most of the available techniques that purport to discriminate between the two only utilize mean RT and differ mainly in the independent variable that is

emphasized. This chapter is devoted to the self-terminating vs. exhaustive processing issue. We shall investigate techniques that vary the number of nontargets, the number of targets, and the serial position of any target within the display. (As usual, the *serial* in *serial position* has nothing to do with the parallel vs. serial issue but is simply an historically established name for the spatial or temporal position of an element.) Then we shall examine RT variances to see if they can aid self-terminating–exhaustive identifiability, and finally we shall briefly consider some distributional techniques. It is hoped that our efforts, by reducing the total uncertainty, will not only increase discriminability on this dimension, but in so doing will indirectly increase identifiability elsewhere in the system.

It is possible, of course, to conceive of systems where processing is neither self-terminating nor exhaustive. For example, following processing of the critical subset it may be that the system begins to slowly wind itself down, but before it terminates, some portion of the remaining stimulus pattern is completed. Other examples, such as Nickerson's (1966) guessing strategy, are discussed by Taylor (1975, 1976c). Such hybrid systems must receive little attention in this chapter.

A convenient experimental foundation upon which to construct paradigms that effectively discriminate between self-terminating and exhaustive processing are the memory-scanning (ET) and visual search (LT) tasks discussed in the preceding chapter. In their unaltered form their power is limited, and for this reason sizable elaborations are necessary. For instance, we shall examine the effects of generalizing these paradigms by varying the number of critical or target elements in displays of any given size. This can make the paradigm much more effective, but at the same time, since we will be manipulating several variables simultaneously, it can make the discussion somewhat more confusing. This problem is easily solved by using the following notational convenience.

Consider a trial of a typical search task (i.e., ET or LT) with n elements in the stimulus array and in which j of the n elements are targets. For many of our investigations, these two numbers will convey all the relevant information about that trial. For instance, by subtracting j from n we can determine the number of nontargets in the array. Thus, we will adopt the convenient notation (j, n) to describe experimental conditions. For instance, $(0, n)$ means that n elements are presented, none of which is a target (a standard nontarget trial); $(1, 1)$ means one element is presented and it is a target; and so on.

In addition to a close scrutiny of the number of target elements we will also find occasion to consider the serial position of an element in the display. Given a linear horizontal stimulus array of size n, every item within the array can be uniquely designated by its serial position. For instance, the element in the ith serial position is characterized by the fact that there are $i-1$ items to its left and $n-i$ items to its right. As we will be concerned primarily with probabilistic systems, we can now define the random variable $\mathbf{T}^{+}_{i,n}$ ($\mathbf{T}^{-}_{i,n}$) as

the time it takes the serial system to process (i.e., compare) the target (nontarget) element that is in the ith serial position in a display of n elements. As usual, when processing is parallel we will instead use the symbol $\mathbf{T}$ when referring to actual processing times and reserve $\mathbf{T}$ for instances in which we are interested in intercompletion times.

Note that this notation allows $\mathbf{T}$ (and equivalently $\mathbf{T}$) to depend on both serial position and display size, which is implied by the statement

$$\mathbf{T}^{\pm}_{i+1,n} \neq \mathbf{T}^{\pm}_{i,n} \neq \mathbf{T}^{\pm}_{i,n+1}$$

where $\mathbf{T}^{\pm}$ means $\mathbf{T}^{+}$ or $\mathbf{T}^{-}$. The "$\neq$" is meant to imply a lack of equivalence in probability distribution. Often it will prove advantageous to place some restrictions on $\mathbf{T}$ to simplify notation and reduce free parameters. One frequent restriction will be to assume that processing time depends at most on display size (i.e., it is independent of serial position). Such a restriction implies that processing times for all n elements in the display come from the same distribution regardless of serial position and will be written as

$$\mathbf{T}^{\pm}_{i,n} \equiv \mathbf{T}^{\pm}_{i+1,n} \equiv \mathbf{T}^{\pm}_{\cdot n} \quad \text{for all } i<n$$

A dot in the subscript is meant to imply invariance – in the above case, invariance across serial positions. Thus, $\mathbf{T}^{\pm}_{i\cdot}$ implies that processing time depends at most on serial position and is independent of display size.

Each point on a target-present or target-absent curve is the mean RT to all possible serial position placements of the targets, conditioned on display size. Characterization of these points thus need not include serial position information. Therefore, let $\bar{T}_{(j,n)}$ be the mean processing time for a display of size n when it contains j targets. Using this notation, target-absent curves, for instance, will be constructed from the points $\bar{T}_{(0,1)}, \bar{T}_{(0,2)}, \ldots, \bar{T}_{(0,n)}$. The term *rate* will sometimes be employed in a general way to refer to speed of processing. In most cases, it may be precisely defined as $1/E(\mathbf{T})$, that is, as the reciprocal of the expected processing time.

We are now sufficiently prepared for a more analytical investigation of large classes of exhaustive and self-terminating models and testing paradigms that purport to discriminate between them. A more detailed treatment of these issues can be found in Ashby (1976).

Testing paradigms

Testing for equal slopes

Typically the parallel (i.e., equal-sloped) target-present and target-absent curves that are so frequently found in the standard ET and LT paradigms are cited as evidence of an exhaustive search (see Chapter 6). However, as is becoming well known, such conclusions are based on extremely strong and possibly untenable assumptions. As we saw in the preceding

chapter, large classes of serial *and* parallel self-terminating models can predict equal slopes for target-present and target-absent curves (e.g., Townsend & Roos 1973).

As yet another example of this fact, consider a serial self-terminating model and conditions $(0,n)$, $(0,n+1)$ and $(1,n)$, $(1,n+1)$. If we assume $\mathbf{T}^{\pm}$ depends at most on display size, then the model predicts

$$\bar{T}_{(0,n)} = nE(\mathbf{T}^{-}_{\cdot n})$$

$$\bar{T}_{(1,n)} = \frac{1}{n}\sum_{j=1}^{n}[E(\mathbf{T}^{+}_{\cdot n}) + (j-1)E(\mathbf{T}^{-}_{\cdot n})]$$

$$= E(\mathbf{T}^{+}_{\cdot n}) + \frac{n-1}{2}E(\mathbf{T}^{-}_{\cdot n})$$

and

$$\bar{T}_{(0,n+1)} = (n+1)E(\mathbf{T}^{-}_{\cdot n+1})$$

$$\bar{T}_{(1,n+1)} = E(\mathbf{T}^{+}_{\cdot n+1}) + \frac{n}{2}E(\mathbf{T}^{-}_{\cdot n+1})$$

Suppose we now restrict ourselves to an even smaller subset of models by assuming that $\mathbf{T}^{-}_{\cdot n}$ is equivalent to $\mathbf{T}^{-}$ for all values of n. Such a restriction implies that all nontarget-processing times are identically distributed random variables and therefore have the same expected value. The following result specifies the conditions under which this class of serial self-terminating models predict parallel target-present and absent curves.

Proposition 7.1: The serial self-terminating model with random variables $\mathbf{T}^{-}$ and $\mathbf{T}^{+}_{\cdot n}$ predicts parallel target-present and target-absent curves if

$$E(\mathbf{T}^{+}_{\cdot n+1}) = E(\mathbf{T}^{+}_{\cdot 1}) + \frac{n-1}{2}E(\mathbf{T}^{-})$$

Proof: Parallel curves result if $\bar{T}_{(0,n+1)} - \bar{T}_{(1,n+1)} = \bar{T}_{(0,n)} - \bar{T}_{(1,n)}$. Under the parameter restrictions of the model this condition becomes

$$(n+1)E(\mathbf{T}^{-}) - \left[E(\mathbf{T}^{+}_{\cdot n+1}) + \frac{n}{2}E(\mathbf{T}^{-})\right]$$

$$= nE(\mathbf{T}^{-}) - \left[E(\mathbf{T}^{+}_{\cdot n}) + \frac{n-1}{2}E(\mathbf{T}^{-})\right]$$

which simplifies to

$$\frac{n+2}{2}E(\mathbf{T}^{-}) - E(\mathbf{T}^{+}_{\cdot n+1}) = \frac{n+1}{2}E(\mathbf{T}^{-}) - E(\mathbf{T}^{+}_{\cdot n})$$

This expression can be rewritten as $E(\mathbf{T}^{+}_{\cdot n+1}) = E(\mathbf{T}^{+}_{\cdot n}) + \frac{1}{2}E(\mathbf{T}^{-})$. The result is achieved through iteration. □

Thus, a serial, self-terminating model predicts equal slopes if target-processing time increases with set size in a manner satisfying the Proposition 7.1 restriction. Another, possibly more appealing model that is similar to this one has target-processing duration depend on the serial position but not on display size. Equal slopes can result if target duration increases with serial position. Although this model predicts a larger difference between adjacent serial positions than the across-set size difference of Proposition 7.1, it does have the advantage of not having to explain how the system knows set size a priori.

This simple illustration reinforces the conclusion that unless one has some strong empirical justifications (for instance, knowledge of processing rate), no conclusion can be reached about whether processing is self-terminating or exhaustive on the basis of equal slopes of target-present and target-absent curves. More generality and detail on this issue can be found in Chapter 6. The upshot of all this is that other testing strategies should be sought.

It is necessary first to digress for a moment and restate an assumption essential to these types of analyses – that is, that the base time (i.e., residual or extra "unstudied latencies"), or that time between stimulus exposure and response completion not including comparison time, is invariant (at least at the level of the means) across display sizes. More precisely, we assume $\overline{\mathrm{RT}}_{(i,n)} - \bar{T}_{(i,n)} = \overline{\mathrm{RT}}_{(i,n+k)} - \bar{T}_{(i,n+k)}$ for any k, where $\overline{\mathrm{RT}}_{(i,n)}$ includes mean base time and $\bar{T}_{(i,n)}$ does not. This assumption allows the inference that any change in mean RT caused by increasing display size occurs entirely within the $\bar{T}_{(i,n)}$ or comparison time. Thus when writing out model predictions, a residual base time term need not be included, for it is assumed a constant. This does not imply that base time is independent of task, but only that it is independent of increases in task length.

Varying the number of redundant targets

The type of paradigm discussed above attempts to test for processing strategy (self-terminating or exhaustive) by varying display size while holding the number of targets, when present, a constant. Such a paradigm may be conveniently described as (c, y), where $y = 1, 2, \ldots, n$ and c is a constant that usually equals 1. This type of paradigm was adopted by Sternberg (1966) and later used by many others. As we have seen, it requires very strong assumptions before any discrimination of search strategy is possible.

The converse to this approach has the experimenter varying the number of targets while holding display size constant. This type of paradigm may be described as (x, n), where $x = 0, 1, \ldots, n$ and n is a constant that is usually greater than or equal to 3. This approach has also been very popular (e.g., Estes & Taylor 1964; Bjork & Estes 1971; van der Heijden 1975). The argument usually runs as follows: If mean RT is a decreasing function of the number of redundant targets, then a self-terminating search is supported, since the more targets in the display the fewer need be scanned before a target is completed. Conversely, if mean RT is a flat or increasing function of the number of redundant targets, then the system must always be processing all of the stimulus array regardless of how many targets there are and hence processing is exhaustive.

Van der Heijden and Menckenberg (1974) and van der Heijden (1975) used such a paradigm in a visual search task (i.e., an ET task), and on the basis of their results argued for a self-terminating process. That is, they generally found mean RT to be a decreasing function of the number of redundant targets. This has become the classical result for these conditions (e.g., Bjork & Estes 1971).

Before we accept their conclusions, we should determine the severity of their assumptions by investigating under what conditions, if any, various classes of exhaustive models can predict mean RT to be a decreasing function of the number of redundant targets. We will start with serial models and then proceed to parallel models.

The paradigm appears to have been designed with serial models in mind. At any rate, it will become clear that many serial self-terminating models can predict the desired decrease in mean RT as the number of targets increases. We therefore begin by considering serial exhaustive models. Van der Heijden employed the conditions $(0, 3)$, $(1, 3)$, $(2, 3)$, and $(3, 3)$, so we will concentrate on these in this section, but extensions to the general case are self-evident.

Proposition 7.2: Serial exhaustive models predict mean RT to be a flat or increasing function of the number of targets in a display of fixed size only if the average nontarget mean processing time is less than or equal to the average target mean time. If mean target times are less than mean nontarget times, then a decreasing function is predicted.

Proof: The serial exhaustive mean processing time predictions for the conditions of interest are

$$\bar{T}_{(0,3)} = \sum_{i=1}^{3} E(\mathbf{T}_{i,3}^{-}) = 3\left[\tfrac{1}{3}\sum_{i=1}^{3} E(\mathbf{T}_{i,3}^{-})\right]$$

$$\bar{T}_{(1,3)} = \left[\tfrac{1}{3}\sum_{i=1}^{3} E(\mathbf{T}_{i,3}^{+})\right] + 2\left[\tfrac{1}{3}\sum_{i=1}^{3} E(\mathbf{T}_{i,3}^{-})\right]$$

$$\bar{T}_{(2,3)} = 2\left[\tfrac{1}{3}\sum_{i=1}^{3} E(\mathbf{T}_{i,3}^{+})\right] + \left[\tfrac{1}{3}\sum_{i=1}^{3} E(\mathbf{T}_{i,3}^{-})\right]$$

$$\bar{T}_{(3,3)} = 3\left[\tfrac{1}{3}\sum_{i=1}^{3} E(\mathbf{T}_{i,3}^{+})\right]$$

Clearly whether these mean processing times are increasing, flat, or decreasing functions of the number of targets in the display depends on whether the average nontarget mean processing time, $\frac{1}{3}\Sigma_{i=1}^{3} E(\mathbf{T}_{i,3}^{-})$, is less than, equal

to, or greater than the average target mean processing time $\frac{1}{3}\sum_{i=1}^{3} E(\mathbf{T}_{i,3}^{+})$. Generalization to other values of n is immediate. □

It does not require an especially vivid imagination to conceive of instances in which target elements might be processed faster than nontarget elements. In fact, it has been argued by some that this is the reason that "same" responses are so consistently found to be faster than "different" responses in same-different tasks (e.g., Bamber 1969; Krueger 1978; see also Chapter 6). At any rate, the severity of the assumption that this testing paradigm makes – that nontarget-processing rate is greater than or equal to target-processing rate – is at present empirically undetermined and hence its acceptance should not be without careful consideration. Should it be deemed untenable, then the value of this type paradigm should be reconsidered.

Next we examine the parallel model predictions. For both the self-terminating and exhaustive cases we will assume only one target rate and one nontarget rate per display size. Even simplifying to this level, however, does not make computation of the expected processing times an easy matter. With parallel processing we need to compute the expected value of the maximum (in the case of an exhaustive search) or the minimum (in the case of self-termination) of a set of individual element processing times. As it turns out, the easiest way to perform these computations is to use the fact that for any random variable $\mathbf{T}$, which can take on only nonnegative values, the expected value of $\mathbf{T}$ can be found from[1]

$$E(\mathbf{T}) = \int_0^\infty [1 - G(\tau)]\, d\tau$$

where $G(\tau)$ is the distribution function associated with $\mathbf{T}$. Thus, any expected processing time is equal to the integral of the survivor function from zero to infinity.

This method of computing expected values will lead to natural orderings on the individual element processing time distribution functions rather than on the means. For instance, we might discover that a certain mean RT curve increases if the individual *nontarget* element processing time distribution function is always greater than the *target* element processing time distribution function, $G_{\cdot n}^{-}(\tau) > G_{\cdot n}^{+}(\tau)$ for all $\tau > 0$. We shall investigate distribution function orderings in some detail in the next chapter, but for now it will suffice to know that such an ordering implies a concomitant ordering on the means but

[1] This expression can be verified by integrating by parts. Let $u = 1 - G(t)$ and $dv = dt$ so that $du = -g(t)\, dt$ and $v = t$. Then

$$\int_0^\infty [1 - G(t)]\, dt = [1 - G(t)]\, t\Big|_0^\infty + \int_0^\infty t g(t)\, dt$$

$$= \int_0^\infty t g(t)\, dt$$

which is the usual definition of expectation.

not vice versa. Thus the above distribution function ordering implies that $E(\mathbf{T}_{\cdot n}^{-}) > E(\mathbf{T}_{\cdot n}^{+})$, but this mean ordering does not imply the distribution function ordering. Our results for parallel models, then, will be stronger when stated in terms of the distribution functions.

Proposition 7.3: Independent parallel self-terminating models without serial position effects always predict $\bar{T}_{(1,3)} > \bar{T}_{(2,3)} > \bar{T}_{(3,3)}$ and they predict $\bar{T}_{(0,3)} > \bar{T}_{(1,3)} > \bar{T}_{(2,3)} > \bar{T}_{(3,3)}$ if $G_{\cdot 3}^{+}(\tau) > [G_{\cdot 3}^{-}(\tau)]^3$ for all $\tau > 0$.

Proof: Parallel self-terminating predictions are

$$\bar{T}_{(0,3)} = E[\max(\mathbf{T}_{\cdot 3}^{-}, \mathbf{T}_{\cdot 3}^{-}, \mathbf{T}_{\cdot 3}^{-})], \quad \bar{T}_{(1,3)} = E(\mathbf{T}_{\cdot 3}^{+}),$$

$$\bar{T}_{(2,3)} = E[\min(\mathbf{T}_{\cdot 3}^{+}, \mathbf{T}_{\cdot 3}^{+})], \quad \text{and} \quad \bar{T}_{(3,3)} = E[\min(\mathbf{T}_{\cdot 3}^{+}, \mathbf{T}_{\cdot 3}^{+}, \mathbf{T}_{\cdot 3}^{+})]$$

Assuming an independent (in total completion times, which implies no capacity reallocation) stochastic system (i.e., nonzero processing time variances), then it is clear that $\bar{T}_{(1,3)} > \bar{T}_{(2,3)} > \bar{T}_{(3,3)}$, with the magnitude of the difference depending on the size of the variance of the completion time distributions.

In addition, $\bar{T}_{(0,3)} > \bar{T}_{(1,3)}$ if and only if

$$\int_0^\infty \{1 - [G_{\cdot 3}^{-}(\tau)]^3\}\, d\tau > \int_0^\infty [1 - G_{\cdot 3}^{+}(\tau)]\, d\tau$$

which is equivalent to the condition that

$$\int_0^\infty \{G_{\cdot 3}^{+}(\tau) - [G_{\cdot 3}^{-}(\tau)]^3\}\, d\tau > 0$$

This inequality is obviously satisfied under the conditions of the proposition. Generalization to arbitrary n is easy and left to the reader. □

Most parallel self-terminating models thus predict a steady decrease in mean RT as the number of targets is increased. We have to be a little more careful if the point $\bar{T}_{(0,3)}$ is included in our analyses, but even if it is, most parallel self-terminating models still predict a decrease. In fact, $\bar{T}_{(0,3)} < \bar{T}_{(1,3)}$ only if nontargets are processed *much* faster than targets. For instance, nontargets are faster if $G_{\cdot 3}^{-}(\tau) > G_{\cdot 3}^{+}(\tau)$ for all $\tau > 0$, but the conditions of the proposition require $G_{\cdot 3}^{-}(\tau) \geq [G_{\cdot 3}^{-}(\tau)]^3 > G_{\cdot 3}^{+}(\tau)$ for all $\tau > 0$, which can only be satisfied if target elements are processed much slower than nontargets.

Finally we examine the predictions of parallel exhaustive models.

Proposition 7.4: Independent parallel exhaustive models without serial position effects predict mean RT to be a flat or increasing function of the number of targets in a display of fixed size if $G_{\cdot n}^{-}(\tau) \geq G_{\cdot n}^{+}(\tau)$ for all $\tau > 0$. A decreasing function is predicted if $G_{\cdot n}^{+}(\tau) > G_{\cdot n}^{-}(\tau)$ for all $\tau > 0$.

Proof: Under the conditions of the theorem,

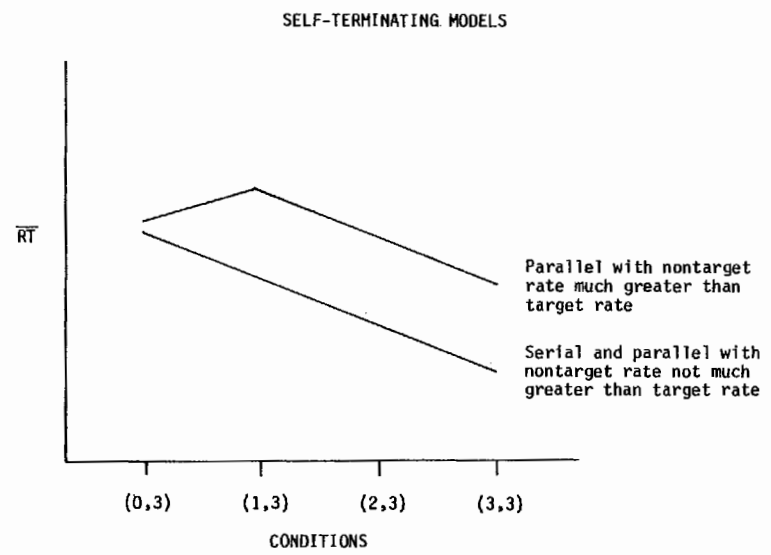


Fig. 7.1. Mean RT results expected of self-terminating models in the redundant targets paradigm.

$$\bar{T}_{(j+1,n)} - \bar{T}_{(j,n)} = \int_0^\infty \{1 - [G^+_{\cdot n}(\tau)]^{j+1}[G^-_{\cdot n}(\tau)]^{n-j-1}\}\, d\tau$$

$$- \int_0^\infty \{1 - [G^+_{\cdot n}(\tau)]^{j}[G^-_{\cdot n}(\tau)]^{n-j}\}\, d\tau$$

$$= \int_0^\infty \{[G^+_{\cdot n}(\tau)]^{j}[G^-_{\cdot n}(\tau)]^{n-j} - [G^+_{\cdot n}(\tau)]^{j+1}[G^-_{\cdot n}(\tau)]^{n-j-1}\}\, d\tau$$

$$= \int_0^\infty [G^+_{\cdot n}(\tau)]^{j}[G^-_{\cdot n}(\tau)]^{n-j-1}[G^-_{\cdot n}(\tau) - G^+_{\cdot n}(\tau)]\, d\tau$$

Clearly if $G^-_{\cdot n}(\tau) \geqslant G^+_{\cdot n}(\tau)$ for all $\tau > 0$, then $\bar{T}_{(j+1,n)} - \bar{T}_{(j,n)} \geqslant 0$ for all $j+1 \leqslant n$, and thus the curve is flat or increasing. Equally clear is that if $G^-_{\cdot n}(\tau) < G^+_{\cdot n}(\tau)$ for all $\tau > 0$, then the curve decreases. □

As was the case with self-termination, parallel predictions are similar to the serial model predictions when search is exhaustive. Whether the mean RT curve increases, remains flat, or decreases is again heavily dependent on the difference between the processing times of target and nontarget elements. The major problems arise with exhaustive models. Most self-terminating models predict a mean RT decrease as the number of replicas of the target element is increased. Exhaustive models, though, can predict virtually any kind of function. If nontargets are processed faster than targets, an increasing function is predicted. If nontargets are processed slower, a decreasing

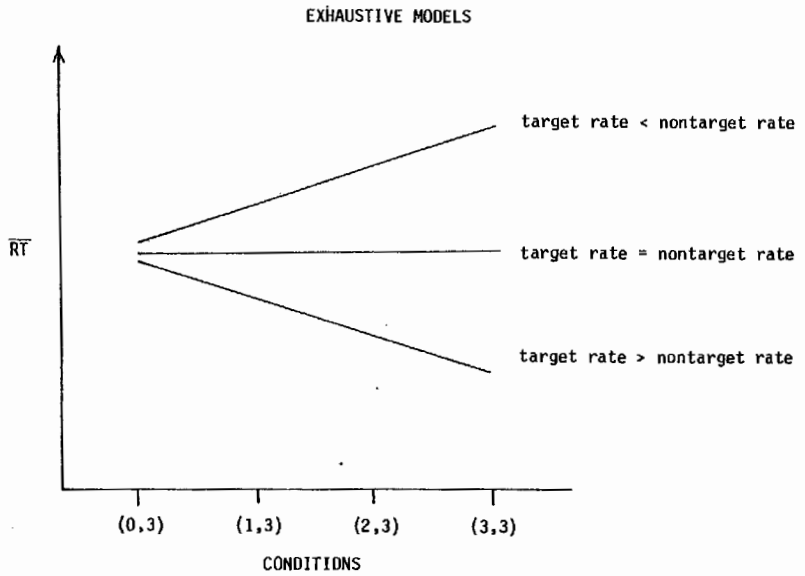


Fig. 7.2. Mean RT results expected of exhaustive models in the redundant targets paradigm.

The main results of this section are summarized in Fig. 7.1 for self-terminating models and in Fig. 7.2 for exhaustive models. It can be seen there that a reasonably accurate conclusion can be drawn about whether search is self-terminating or exhaustive only when an increasing mean RT function is found. In the case of a flat function an exhaustive search is favored, but it may not be possible to rule out a parallel self-terminating model in which targets are processed so slowly that $[G^-_{\cdot 3}(\tau)]^3 > G^+_{\cdot 3}(\tau)$ for some $\tau > 0$. Finally, no firm conclusions can be drawn if a decreasing function is found and, unfortunately, this is the empirical result most frequently reported.

To make matters worse, it appears there may be another phenomenon, having nothing to do with search strategy, that may contribute to decreasing mean RT curves in these conditions. Same–different search tasks are often characterized by very fast "all-same" responses, and since these types of conditions demand an exhaustive search, several researchers (Bamber 1972; Taylor 1975) have suggested that a different, more holistic processing mechanism might operate on these trials. It may be that (3, 3) trials elicit very fast "yes" responses in a manner analogous to "all-same" trials and that the corresponding depression in mean RT at this point causes the curves to spuriously appear to support a self-terminating explanation. This type of "gestalt" perceptual response has been noted in many empirical contexts.

In the paradigm's defense, it does have one big advantage over most of its competitors. Since it holds display size constant, the load on the system is

Capacity limitations can easily obscure identifiability of other unrelated issues, and thus if one suspects limited capacity, then this type of paradigm represents a reasonable way of dealing with this parameter.

It is important to realize that this testing paradigm is not being singled out because it apparently fails in its quest, but rather that it is being discussed because it is one of the more promising paradigms to appear in the literature. Certainly it is superior to the more prevalent technique of testing for a 2:1 slope ratio of the target-absent curve to the target-present curve. Such studies not only assume equal processing rates for targets and nontargets and that rate is independent of serial position and display size, but also that processing is serial. The major assumption of the redundant target paradigm is that nontarget-processing rate is greater than or equal to target rate. Thus in this case, target–nontarget rate differences can obscure identifiability, but interestingly enough, at times they have been utilized to promote it. For instance, Townsend (1976a) proved a number of theorems suggesting that with exponential models just such a rate difference ensures that identifiability is at least possible (see Chapter 13).

Keeping the expected number of items processed a constant

Both paradigms discussed above attempt to distinguish between self-terminating and exhaustive search by assuming self-termination and trying to vary the mean number of elements processed before a target is completed. The first paradigm tries to increase this mean by adding nontargets to the display, whereas van der Heijden and others tried to decrease it by replacing nontargets in the display with targets. Ashby (1976) suggested a new approach, that of holding the mean constant while varying display size.

With one target in a display of y elements, a serial, self-terminating model predicts that an average of $(y+1)/2$ elements must be processed before the target is completed. In general, with x targets, the number of elements to be processed before the first target is completed has an inverse hypergeometric distribution with parameters y, x, and 1, and mean equal to $(y+1)/(x+1)$ (Patil & Joshi 1968). Ashby suggested varying both the number of targets, x, and display size, y, in such a way that $(y+1)/(x+1)$ is constant. For instance, if we choose conditions (1, 3), (2, 5), and (3, 7), then the expected number of items processed before the first target is completed will be two for all three conditions. Intuitively one expects an exhaustive model to predict an increase in mean RT with these display size increases and a self-terminating model to predict no effect of display size on RT. Such predictions can be verified, and in the process the assumptions necessary for identifiability can be discovered.

The predictions of serial self-terminating models with equiprobable processing paths and rate free to depend on both serial position and set size are very tedious with n as large as 7 [i.e., (3, 7)]. However, we should keep in mind that this paradigm, like the ones discussed earlier in this chapter, was constructed with these serial models in mind; our intuitions will in this case be reliable. If all target elements are processed at the same rate and if all nontarget elements are processed at the same rate, then serial self-terminating models predict $\bar{T}_{(1,3)}=\bar{T}_{(2,5)}=\bar{T}_{(3,7)}=E(\mathbf{T}^-)+E(\mathbf{T}^+)$ and therefore the curve is flat. In general, as long as $E(\mathbf{T}^{\pm})$ does not increase with set size, then the curve determined by $\bar{T}_{(1,3)}$, $\bar{T}_{(2,5)}$, and $\bar{T}_{(3,7)}$ will be nonincreasing. This seems to be a reasonably weak restriction. Serial self-terminating systems will, given the above three experimental conditions, on the average always process two elements (one target and one nontarget), regardless of the display size. Thus, the only way mean RT can increase is if the processing time of, say, the second element takes longer when there are more elements in the display. But this implies a priori knowledge by the system of the display size (or processing load) on any given trial. In a parallel system such knowledge is a natural consequence of the system's structure, and we attribute this effect of display size on processing rate to capacity limitations. But in serial systems this notion is less intuitive, especially in visual search (ET) tasks. Instead, we might imagine a limited capacity serial system as one that slows down as more and more elements are processed. For instance, we might imagine that the processing rate of ensuing elements decreases as more elements are completed. Under this interpretation, the second element is processed more slowly than the first element, the third element is processed more slowly than the second, etc. This type of limited capacity model will still predict $\bar{T}_{(1,3)}=\bar{T}_{(2,5)}=\bar{T}_{(3,7)}$, as we expect of self-terminating models.

In serial exhaustive models $\bar{T}_{(1,3)}=E(\mathbf{T}_3^+)+2E(\mathbf{T}_3^-)$, $\bar{T}_{(2,5)}=2E(\mathbf{T}_5^+)+3E(\mathbf{T}_5^-)$, and $\bar{T}_{(3,7)}=3E(\mathbf{T}_7^+)+4E(\mathbf{T}_7^-)$, where $E(\mathbf{T}_n^+)$ is the mean of the target processing times when the display size is n, that is,

$$E(\mathbf{T}_n^+)=\frac{1}{n}\sum_{i=1}^{n}E(\mathbf{T}_{i,n}^+)$$

Similarly, $E(\mathbf{T}_n^-)$ is the mean of the nontarget times. Except in cases of extreme supercapacity, this mean RT curve will rise sharply with increases in display size; in other words, all limited capacity, all unlimited capacity, and many supercapacity serial exhaustive models predict $\bar{T}_{(1,3)}<\bar{T}_{(2,5)}<\bar{T}_{(3,7)}$. The magnitude of this increase will, of course, depend on $E(\mathbf{T}_n^+)$ and $E(\mathbf{T}_n^-)$. Nevertheless, it appears safe to say that with regard to serial models at least, the paradigm appears to be a potentially powerful discriminator between self-terminating and exhaustive search strategies.

We next examine independent parallel models. Since capacity is clearly an important factor in this paradigm, we will assume that processing rate can depend on display size and element identity (target or nontarget). As before, the results that follow will place orderings directly on the individual element completion time distribution functions. As a consequence, we need definitions of capacity at this level. The construction of such definitions forms a major part of the next chapter, and so a thorough discussion of this problem is delayed until then. Serial position effects significantly complicate these definitions. Without them, the definitions are exceedingly simple. Since we are not assuming serial position differences here anyway, the simplified definitions will suffice for our present purposes.

To determine the capacity of any given system, one must compare the individual element completion times as the processing load changes. If they increase, we say the system is limited capacity; if they decrease, it is supercapacity; and if they remain the same, we say the system is unlimited capacity. Basically we will do the same thing here, only we will define *processing time* in terms of the distribution functions rather than, say, the means. Thus, we will call a system *limited* capacity if

$$G_{\cdot n-1}(\tau) > G_{\cdot n}(\tau) \quad \text{for all } \tau > 0$$

We will say a system is *super*capacity if

$$G_{\cdot n-1}(\tau) < G_{\cdot n}(\tau) \quad \text{for all } \tau > 0$$

and otherwise that it is *unlimited* capacity.

Proposition 7.5: Independent parallel self-terminating models without serial position effects predict $\bar{T}_{(1,3)} \geq \bar{T}_{(2,5)} \geq \bar{T}_{(3,7)}$ except when capacity is *extremely* limited.

Proof: Parallel self-terminating models predict

$$\bar{T}_{(1,3)} = E(\mathbf{T}^{+}_{\cdot 3}), \quad \bar{T}_{(2,5)} = E[\min(\mathbf{T}^{+}_{\cdot 5}, \mathbf{T}^{+}_{\cdot 5})], \quad \text{and}$$
$$\bar{T}_{(3,7)} = E[\min(\mathbf{T}^{+}_{\cdot 7}, \mathbf{T}^{+}_{\cdot 7}, \mathbf{T}^{+}_{\cdot 7})]$$

The curve is nonincreasing when

$$\int_0^\infty \bar{G}^{+}_{\cdot 3}(\tau)\, d\tau \geq \int_0^\infty [\bar{G}^{+}_{\cdot 5}(\tau)]^2\, d\tau \geq \int_0^\infty [\bar{G}^{+}_{\cdot 7}(\tau)]^3\, d\tau$$

because, for example, the survivor function of $\min(\mathbf{T}^{+}_{\cdot 5}, \mathbf{T}^{+}_{\cdot 5})$ is $[\bar{G}^{+}_{\cdot 5}(\tau)]^2$. A sufficient condition for the curve to be nonincreasing is therefore

$$\bar{G}^{+}_{\cdot 3}(\tau) \geq [\bar{G}^{+}_{\cdot 5}(\tau)]^2 \geq [\bar{G}^{+}_{\cdot 7}(\tau)]^3 \quad \text{for all } \tau > 0 \tag{7.1}$$

In the next chapter, we will see that in the case of the minimum completion time, *fixed capacity* models are those for which

$$\bar{G}_{\cdot 1}(\tau) = [\bar{G}_{\cdot n}(\tau)]^n \quad \text{for all } \tau > 0$$

This definition of fixed capacity implies that

$$[\bar{G}_{\cdot 3}(\tau)]^3 = [\bar{G}_{\cdot 5}(\tau)]^5 = [\bar{G}_{\cdot 7}(\tau)]^7 \quad \text{for all } \tau > 0$$

Models even more limited in capacity are called *extremely limited capacity* models. Comparing these constraints with Eq. 7.1, it can be seen that fixed capacity models as well as some extremely limited capacity models predict a nonincreasing curve. □

Many limited capacity models predict mean RT to be a decreasing function of display size under these experimental conditions. Included in this group are fixed capacity models. We encountered these in Chapter 4 in the context

of exponential processing. When the individual element completion times are exponentially distributed, then fixed capacity implies that $v_n = (1/n)v_1$. The idea here is that, no matter what the processing load, there is always a fixed supply of capacity to be allocated in some way to the to-be-processed elements. We can think of the total amount of capacity available as v_1. Thus, for every processing load, the individual element processing rates must always sum to v_1. With a load of n elements, each individual rate is $(1/n)v_1$, and so the sum of the n rates is v_1, as required.

Extremely limited capacity models are even more limited in capacity than fixed capacity models. These models place a severe restriction on capacity. For instance, fixed capacity as well as extremely limited capacity independent parallel models predict a positively accelerated target-absent mean RT curve in a standard memory-scanning (LT) or visual search (ET) task (see Chapter 4). As is well known, target-absent curves are usually found to increase (at most) linearly, indicating that capacity may be less limited than fixed capacity independent parallel models propose.

The next result gives the predictions of parallel exhaustive models.

Proposition 7.6: Independent parallel exhaustive models without target-nontarget rate differences and without serial position effects predict $\bar{T}_{(1,3)} < \bar{T}_{(2,5)} < \bar{T}_{(3,7)}$ if

$$[G_3(\tau)]^3 \geq [G_5(\tau)]^5 \geq [G_7(\tau)]^7 \quad \text{for all } \tau > 0$$

with a strict inequality holding for at least one $\tau > 0$.

Proof: An increasing mean RT curve is predicted if $\bar{T}_{(1,3)} < \bar{T}_{(2,5)} < \bar{T}_{(3,7)}$ or, equivalently, if and only if

$$\int_0^\infty \{1 - [G_3(\tau)]^3\}\, d\tau < \int_0^\infty \{1 - [G_5(\tau)]^5\}\, d\tau < \int_0^\infty \{1 - [G_7(\tau)]^7\}\, d\tau$$

Clearly the conditions stated in the proposition are sufficient to guarantee this inequality. □

All limited capacity models satisfy the Proposition 7.6 inequality, and thus they all predict an increasing mean RT function. In fact, many supercapacity models also satisfy the inequality. For instance, suppose

$$G_n(\tau) = [G_1(\tau)]^{1/(n-1)}, \quad \text{for all } \tau > 0$$

Unlimited capacity results when $G_n(\tau) = G_1(\tau)$ for all $\tau > 0$, and so the above model is clearly super in capacity. Even so, it implies

$$[G_3(\tau)]^3 = [G_1(\tau)]^{3/2}, \quad [G_5(\tau)]^5 = [G_1(\tau)]^{5/4}, \quad \text{and} \quad [G_7(\tau)]^7 = [G_1(\tau)]^{7/6}$$

and because

$$[G_1(\tau)]^{3/2} \geq [G_1(\tau)]^{5/4} \geq [G_1(\tau)]^{7/6} \quad \text{for all } \tau > 0$$

this supercapacity model still predicts an increasing mean RT curve.

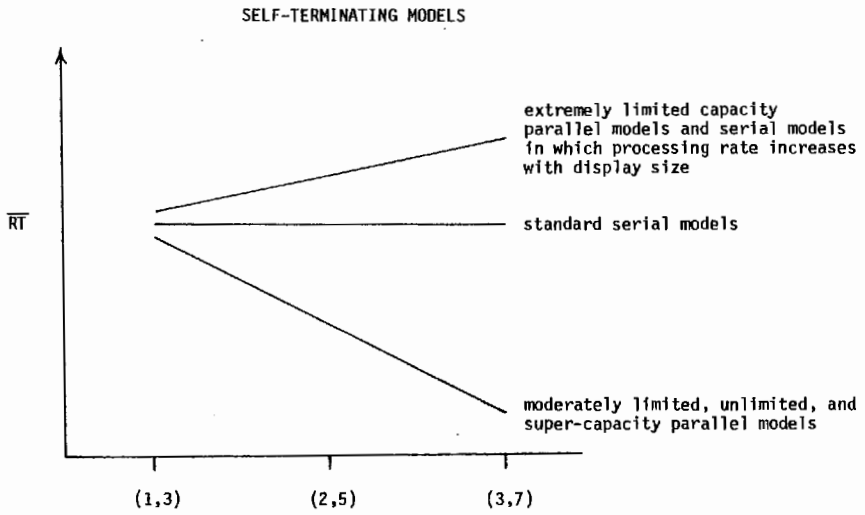


Fig. 7.3. Mean RT results expected of self-terminating models in Ashby's (1976) paradigm.

The boundary case that separates those models predicting an increasing curve from those predicting a decreasing curve is the extremely supercapacity model in which

$$G_n(\tau) = [G_1(\tau)]^{1/n} \quad \text{for all } \tau > 0$$

This model, which is even more super in capacity than our example above, predicts a flat curve; that is, it predicts that $\bar{T}_{(1,3)} = \bar{T}_{(2,5)} = \bar{T}_{(3,7)}$. Models even more super in capacity than this one predict a decreasing mean RT curve.

The paradigm appears to be an effective discriminator between exhaustive and self-terminating search strategies. Figures 7.3 and 7.4 summarize the results of this section. They illustrate that the great majority of models we examined predict that mean processing time is an increasing function across experimental conditions when search is exhaustive but a flat or decreasing function when search is self-terminating. The exceptions tend to be rather unintuitive as models of human information processing. For example, in the class of independent parallel models only those with some inordinate capacity structure fail to make the desired prediction. In the class of serial models, virtually all exhaustive models predict an increase in processing time across conditions, whereas the only self-terminating models making this prediction are forced to postulate the unintuitive notion that processing rate depends upon display size (and increases with it). Thus it appears that this paradigm can analytically discriminate between self-terminating and exhaustive models at a fairly inexpensive cost, assumptionwise. Certainly empirical investigation is called for to substantiate these predictions.

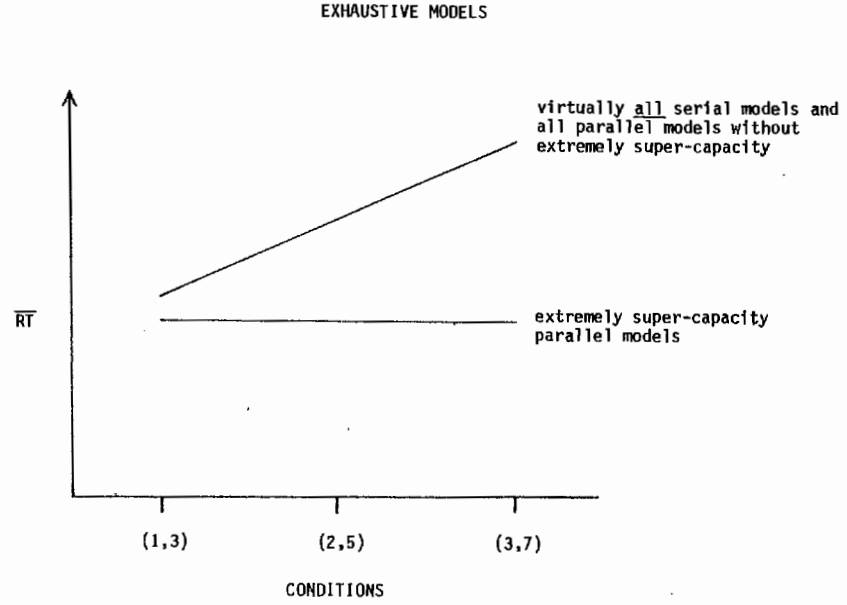


Fig. 7.4. Mean RT results expected of exhaustive models in Ashby's (1976) paradigm.

None of the experimental paradigms we investigated appear to be absolutely foolproof with regard to separating exhaustive from self-terminating processing over all conceivable manipulations of capacity structure and target–nontarget rate differences. Certainly, the strategy of testing for a 2:1 ratio of target-absent to target-present slope in standard ET and LT tasks is by itself unacceptable. Paradigms that vary the number of redundant targets but not display size are an important step in the right direction, but still appear forced into some strong assumptions about target and nontarget processing times. On the other hand, the tack of varying display size, while holding constant the expected number of completions before the first target element is completed, requires much weaker assumptions, and on this basis appears promising, but requires empirical validation.

An empirical application

Even though none of these paradigms is sufficient to guarantee self-terminating–exhaustive identifiability, the collection of the three, taken together, must be a more powerful discriminator than any one of the tests by itself. For instance, whereas there are certain models that each paradigm will misdiagnose (as self-terminating or as exhaustive), these models tend to be different for the three paradigms and thus the intersection of these three sets of models should be very small. We recommend that an empirical application incorporate at least the latter two of these paradigms and perhaps all three.

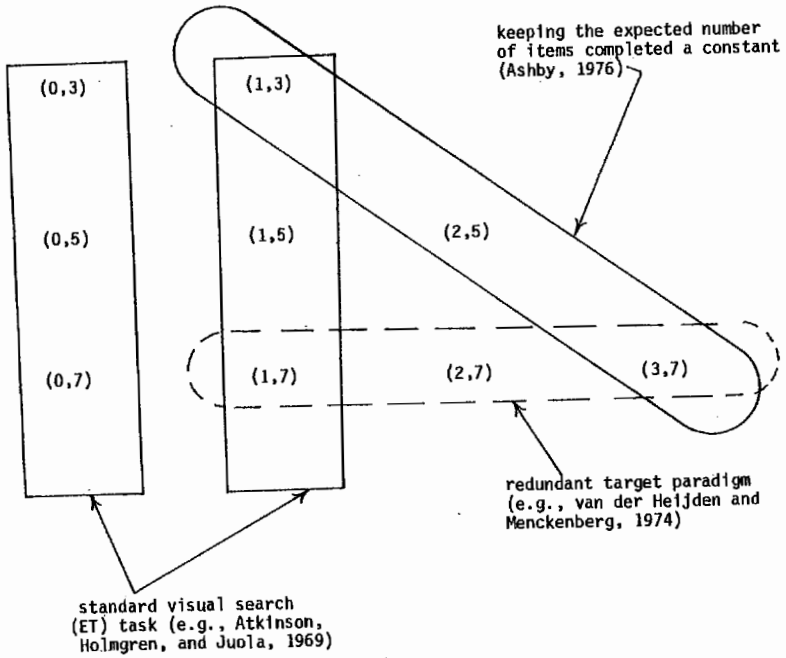


Fig. 7.5. Schematic illustrating the experimental conditions of the pilot study reported in the text.

This was the motivation for a pilot study based on the visual search paradigm that we designed. The experimental conditions employed in the task are illustrated in Fig. 7.5. Recall that a condition (x, y) contains y elements, of which x are targets. As can be seen, with these conditions each of the three tests we examined in this chapter is possible. The standard visual search (ET) task utilizes display sizes of three, five, and seven elements. The redundant target paradigm utilizes conditions (1, 7), (2, 7), and (3, 7). This same test can also be applied, somewhat more tentatively, to conditions (1, 5) and (2, 5). Conditions (0, 7) and (0, 5) could also have been incorporated into the test but were not because of the uppermost curve in Fig. 7.1; parallel self-terminating models with very fast nontarget processing sometimes predict a mean RT increase between conditions $(0, n)$ and $(1, n)$, and thus including the $(0, n)$ condition in the redundant target analysis could actually obscure self-terminating–exhaustive identifiability. Finally, the last paradigm we examined, which holds $(y+1)/(x+1)$ constant, where y is the display size and x is the number of targets, utilizes conditions (1, 3), (2, 5), and (3, 7).

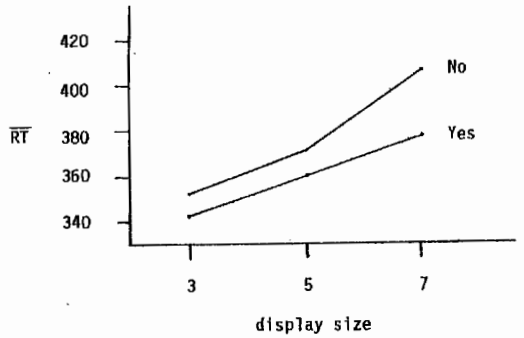
The observers in the study, who were all undergraduates at Purdue University, participated in 5 practice and 15 experimental sessions each lasting approximately one hour. Stimulus elements were capital letters of the English alphabet and were typed (Prestige Elite) on 5″×8″ index cards. A circular

array was used for stimulus presentations in an effort to keep a constant visual angle (1.8°) across conditions (for a more detailed discussion of circular vs. linear arrays see Chapter 6). Stimulus cards were presented to the observers via a Gerbrands T-2B two-field tachistoscope at a luminance of 10 foot-lamberts.

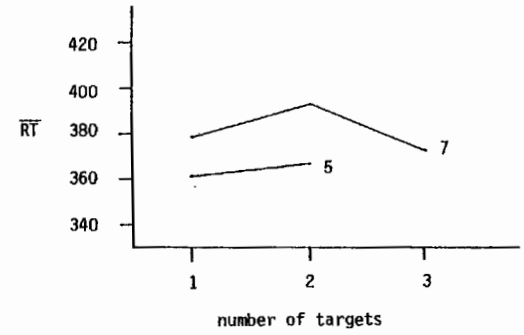
Each experimental session was divided into three blocks, one for each display size. In each block, half the trials required a "no" response ["the target is not in the display," i.e., $(0, y)$ trials] and half required a "yes" response ["the target is in the display," i.e., (x, y) trials, where $1 \leqslant x \leqslant 3$]. The sequence of events on each trial was as follows: The experimenter reads the name of the target item twice; after a short pause the experimenter says, "Ready," and simultaneously presses a button; 100 msec later the circular stimulus array is presented foveally for 150 msec. The observer then presses one of two keys indicating whether the target was contained in the display. Observers were instructed to respond as quickly as possible without sacrificing accuracy for speed.

In the study 255 trials were run for each experimental condition requiring a "yes" response. The requirement that each block contain an equal number of "yes" and "no" trials meant that there were $2 \times 255 = 510$ (0, 5) trials and $3 \times 255 = 765$ (0, 7) trials. Error rates for all observers were less than 10%. Our present analysis deals only with correct RTs. The mean RTs of one observer (error rate = 3.6%) are presented in Fig. 7.6. Beginning with 7.6a we see a slight tendency for the target-absent curve to rise more quickly than the target-present curve. On the basis of these data alone one might argue for self-termination. Before coming to this conclusion however, we should examine Figs. 7.6b and c. The mean RT increase in Fig. 7.6c would be very difficult for a self-terminating model to predict, and the absence of a consistent mean RT decrease in Fig. 7.6b also is evidence against a self-terminating search. Figures 7.6b and c strongly support an exhaustive search. Taken overall, with the preceding theoretical tools in hand, the results are much more compatible with exhaustive processing.

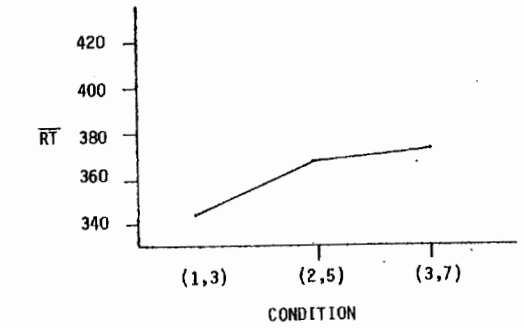
Unfortunately, the analysis of the data from the other observers was not always so straightforward. One of the problems appeared to be differential practice effects. Throughout the course of the experiment, observers had twice as much practice with display size 5 and three times as much practice with display size 7 as they did with display size 3. This may be one of the reasons some observers were actually faster on larger display sizes than on smaller. At any rate, the interpretation of the data of some observers appeared to require a more sophisticated analysis – one that includes some actual model fitting. The purpose of the present chapter is primarily pedagogical. The data presented in Fig. 7.6 fulfill this purpose by illustrating the theoretical results we have developed as well as the power of convergent paradigms. A more complete empirical application of these results, which includes more data and a more sophisticated analysis of these data, will have to be a target of future efforts.



(a)



(b)



(c)

Fig. 7.6. Mean RT data of one observer broken down by experimental paradigm. (a) presents the results from conditions comprising a standard visual search task, (b) from conditions of the redundant targets paradigm, and (c) presents the data from conditions of Ashby's (1976) paradigm.

Serial position curves and identifiability

While the search for the ideal testing paradigm continues, it may prove fruitful to initiate efforts in a new direction. The large identifiability problem that exists with typical mean RT data suggests the need for some additional empirical dimension or anchor from which to test these models. The tack of devising new experimental paradigms certainly has merit, for clever new paradigms can lead to the exclusion of many classes of models from further consideration. This approach, however, is necessarily bounded above by the number of dimensions of the dependent variable. As one utilizes more and more different kinds of data, besides overall mean RT, to identify the underlying processes, identifiability problems should decrease. Logical candidates include error data and serial position curves. Serial position curves seem especially promising, as most viable models make specific quantitative predictions on this dimension.

Serial position curves, that is, mean RT curves conditioned on the position of the target within the display, can be constructed for each display size utilized. Thus for set size n, the serial position curve is determined by n points, where the jth point, $S_{j,n}$, represents mean RT when the target is the jth element in a display of size n. These curves are typically constructed from data in which exactly one target exists in the stimulus array. It is also usually required that the target appear randomly in every position throughout experimentation. This condition, of course, ensures that the curve will contain n points.

The first requirement, that there be only one target in the display, rules out some designs in which RT is the dependent variable, including many of those discussed in the preceding section. In particular, serial position curves can obviously not be constructed for target-absent conditions. This restriction also excludes displays with redundant targets, although in this case it may be possible to construct a multidimensional serial position space. Such a space would require, with j targets, $j+1$ dimensions, one dimension for the dependent variable RT and j dimensions to locate the targets in the display. For instance, with up to two targets a three-dimensional space will suffice where the (x, y) coordinates locate the targets and the z coordinate records RT. Thus the point $(1, 3, 500)$ means that when targets were in serial positions 1 and 3, the mean RT was 500 msec. In this instance, the collection of all these points will generate a serial position *surface*. Although such representations of the data may become difficult to conceptualize when the number of redundant targets is high, they may still be useful, as most models make specific quantitative serial position predictions even when multiple targets exist in the display.

Another assumption frequently made is that the display is presented in a horizontal linear array. As we have seen, however, other types of displays are often used in experimental designs employing RT as a dependent variable

(circular displays, matrix type displays, etc.). Serial position is a nominal scale, and thus in principal there is no reason why "serial" position curves cannot be constructed for any display type as long as a consistent rationale is developed for assigning stimulus elements to serial positions.

Most typically, however, serial position curves are constructed when only one target exists in a linear display (as for example, in the popular ET and LT designs). In this section we will attempt to determine whether such curves can help us discriminate between self-terminating and exhaustive search. To begin, note that for every set of serial position curves there are at least two dimensions of potential interest. First we can condition on set size and examine each serial position curve separately for phenomena such as "recency" or "primacy"; that is, we can examine whether a serial position curve decreases, increases, remains constant, or does something else as the serial position of the target changes. We expect that these effects might be linked to individual element processing rate differences, specifically over the different serial positions and possibly over element identity (e.g., target or nontarget). Another possibility, which is to be considered if processing is serial, is that these effects are caused by the order through which the stimulus list is searched.

Secondly, we can condition on serial position. Here we examine the behavior of the system when the target is in a given serial position as the set size or processing load is changed. We can observe whether increasing the load causes an increase, decrease, or no change in the processing time for any given serial position. We expect these effects to be largely due to the capacity resources of the system, and their allocation to the processing task at hand.

Serial self-terminating models

Let $S_{i,n}$ denote mean processing time when the target is in the ith position in a linear display of n elements. Processing time predictions for a fixed-path (e.g., left-to-right processing) serial self-terminating model are then

$$S_{i,n}=E(\mathbf{T}^-_{1,n})+E(\mathbf{T}^-_{2,n})+\cdots+E(\mathbf{T}^-_{i-1,n})+E(\mathbf{T}^+_{i,n})$$

On the other hand, models that allow the processing path to vary, though more difficult to handle analytically, are also of interest. Here it is expected that serial position data will impose restrictions primarily upon the paths rather than the rates. Predictions of these models are extremely unwieldy, and thus we will limit our investigation to the $n=3$ case and to a particular model that determines path probabilities by using something like Luce's (1959a) choice rule.

Assume every serial position i in the display is assigned a weight b_i such that when the first n positions are available for processing (where $n \geqslant i$), the probability that the element in position i is selected first is equal to $b_i/(\sum_{j=1}^{n} b_j)$. It can be shown that this model predicts

$$S_{1,3}=E(\mathbf{T}^+_{1,3})+\frac{b_2}{b_1+b_2}E(\mathbf{T}^-_{2,3})+\frac{b_3}{b_1+b_3}E(\mathbf{T}^-_{3,3})$$

$$S_{2,3}=E(\mathbf{T}^+_{2,3})+\frac{b_1}{b_1+b_2}E(\mathbf{T}^-_{1,3})+\frac{b_3}{b_2+b_3}E(\mathbf{T}^-_{3,3})$$

and

$$S_{3,3}=E(\mathbf{T}^+_{3,3})+\frac{b_1}{b_1+b_3}E(\mathbf{T}^-_{1,3})+\frac{b_2}{b_2+b_3}E(\mathbf{T}^-_{2,3})$$

Since more parameters are added to the model by letting it vary its processing path, to keep things manageable we will restrict the number of its rate parameters. For this reason assume $\mathbf{T}^+_{i,j}\equiv\mathbf{T}^+$ and $\mathbf{T}^-_{i,j}\equiv\mathbf{T}^-$ for all i and j. This simplifies the above predictions to

$$S_{1,3}=E(\mathbf{T}^+)+\left(\frac{b_2}{b_1+b_2}+\frac{b_3}{b_1+b_3}\right)E(\mathbf{T}^-)$$

$$S_{2,3}=E(\mathbf{T}^+)+\left(\frac{b_1}{b_1+b_2}+\frac{b_3}{b_2+b_3}\right)E(\mathbf{T}^-)$$

and

$$S_{3,3}=E(\mathbf{T}^+)+\left(\frac{b_1}{b_1+b_3}+\frac{b_2}{b_2+b_3}\right)E(\mathbf{T}^-)$$

The success of these two classes of serial self-terminating models in predicting different types of serial position curves will tell us under what conditions the two sets of parameters that characterize these models (i.e., rate parameters and search order parameters) are most effective. For instance, it turns out that when we look at individual curves, that is, when we observe the effects of serial position for a given processing load, the search order parameters are much more important than the individual element rate parameters.[2] Our variable-path model can easily accommodate any type of curve (i.e., increasing, decreasing, or flat) by merely adjusting the magnitudes of b_1, b_2, and b_3.

If these three parameters are all equal, so that all search orders are equally likely, then flat serial position curves are predicted. On the other hand, increasing curves are generated if a left-to-right search is more likely (i.e., if $b_1>b_2>b_3$), whereas decreasing curves indicate a tendency toward right-to-left search ($b_1<b_2<b_3$). On the basis of these results we would expect the fixed-path left-to-right model to easily predict increasing curves without having to resort to its extra rate parameters. This is exactly the case. If we always process in a left-to-right fashion, processing will tend to take longer as the critical element is moved to the right. Even so, this fixed-path model does have enough flexibility to also predict flat and even decreasing curves, although

[2] This proof and all others referred to in this section can be found in Ashby (1976).

such predictions can be achieved only through some rather bizarre assumptions about the different processing rates (for details see Ashby 1976). Certainly, with regard to this prediction, the assumptions of the variable-path model appear much more reasonable.

One plausible possibility is that search order mimics reading behavior. This would suggest that we might find increasing serial position curves in societies where left-to-right reading predominates, at least when serial position effects are found. This prediction does tend to be confirmed (e.g., Atkinson et al. 1969; Townsend & Roos 1973). More rare are studies with observers who naturally read from right to left. Our reading hypothesis predicts that decreasing serial position curves should be more likely with this group.

Interestingly, when we look at the effect of increasing the processing load, given that the target is in some fixed serial position, the importance of the two parameter sets is reversed. The natural prediction of variable-path serial self-terminating models is that processing time for any given serial position of the target will tend to increase as the total display size is increased. A model postulating a strict left-to-right search, however, can accommodate constancy. For instance, if the leftmost element is always processed first, then it is easy to predict that $S_{1,k}=S_{1,n}$ for all $k \leqslant n$.

On the other hand, to predict that processing time is faster on some serial position when *more* elements are in the display may require a large number of processing *rate* parameters. In some cases, the processing path distribution itself might encompass such effects – for example, when, say position 2 is always processed second when $n=2$ but first when $n=3$. Assumptions about search order alone will not suffice in general, however. For instance, consider the case of two displays, one consisting of a single target and the second consisting of a target in position 1 and a nontarget in position 2. In the first display the target must always be processed first, and so even if the target is always processed first in the second display, processing time cannot be faster unless one assumes that the addition of the nontarget element to the display caused the individual element processing rates to change (i.e., decrease). This intuition turns out to be correct. To predict such a speedup overall (or in many situations) we must assume that, on the average, each element in the display is processed faster when the processing load is greater. Again, however we point out that with serial systems, this assumption is not trivial because it implies a priori knowledge by the system of display size. In certain instances, such as standard memory-scanning (LT) tasks, the assumption may not be quite as unreasonable, since the display is always available before search begins; prior processing may indicate the memory size and thus permit rate setting, at least in principle. However, it still requires planning ahead or pacing by the processing system. The system deliberately processes early elements at a slower rate than it needs to, so that it is not too fatigued to process later elements. In visual search (ET) tasks, though, the display is not available until processing begins and so in these cases the assumption seems much

stronger and possibly untenable. Fortunately for this class of models, though, data that would require this assumption are rather infrequently encountered. In the great majority of instances, mean RT to a display with a target in a given serial position increases or remains constant when the display size is increased. These data are easily predicted by the models we have been considering.

Before we move on to examine serial exhaustive models we should first consider some data that we have so far ignored. These are the average target-present and target-absent curves that are so frequently encountered. Although these curves alone cannot discriminate between self-terminating and exhaustive models, observability can only be increased if we simultaneously study target-present, target-absent, and serial position curves.

There are many possible combinations of these different types of curves that we could study. Many of these would tell us very little since they are infrequently encountered empirically. Therefore, we shall restrict our attention to the most frequently encountered possibilities. Specifically, we shall concern ourselves with linear target-present and absent curves with equal slopes. In addition, studies that encouraged serial position effects by asking observers to process stimulus items in the left-to-right fashion (e.g., Townsend & Roos 1973) have obtained serial position curves that are approximately monotonically increasing from left to right, although this effect does not appear as strong as the phenomenon of parallel target-present and target-absent curves. At any rate, we will begin by investigating what restrictions the combination of increasing serial position curves and linear and parallel target-present and target-absent curves impose on serial self-terminating models.

We have already seen several examples of serial self-terminating models that can predict linear parallel target-present and target-absent curves. One way they can do this is by postulating a tendency toward left-to-right search and that the individual target element processing times become longer the more rightward the target is placed. Thus the savings in time produced by the self-terminating nature of search is counterbalanced by the very long individual processing times that occur when the target is placed on the right side of the display.

This model, although somewhat unintuitive, is not completely unbelievable. One could imagine a limited capacity system wherein nontarget elements draw off very little capacity compared to target elements and less capacity is available for the target as it is moved to the right. Such a system could make these predictions. As it turns out, however, we need not bother with developing the model any further, for the serial position curves it predicts immediately falsify this fixed-path model. The problem, which is illustrated in Fig. 7.7, is that the predicted serial position curves have a slope that is twice that of the target-present and target-absent curves (see Ashby 1976 for a proof). Such strong serial position effects are extremely rare.

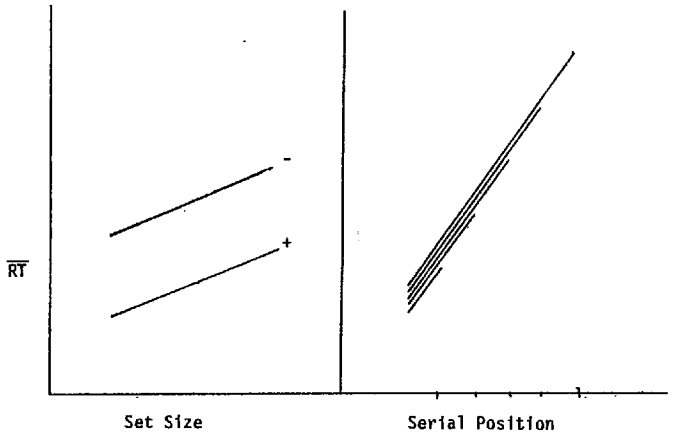


Fig. 7.7. Serial position and set size curves predicted by a serial, self-terminating model with left-to-right search and in which the processing times of target elements are greater the farther to the right the target is placed in the display.

One way to decrease the serial position slope and hence make the model more viable is to allow the target rates to depend on display size rather than on serial position. In other words, suppose $\mathbf{T}^+_{i,n} \equiv \mathbf{T}^+_{i+1,n} \equiv \mathbf{T}^+_{\cdot n}$. Essentially this rotates the serial position curves in Fig. 7.7 about their mean (see Fig. 7.8), thus decreasing their slope but leaving the mean, or in other words, the $\bar{T}_{(1,j)}$ (i.e., the points on the target-present curve), constant. We can quickly verify our intuition about the slopes. Note that now

$$S_{1,n} = E(\mathbf{T}^+_{\cdot n}) = E(\mathbf{T}^-) + [E(\mathbf{T}^+_{\cdot n}) - E(\mathbf{T}^-)]$$

$$S_{2,n} = E(\mathbf{T}^-) + E(\mathbf{T}^+_{\cdot n}) = 2E(\mathbf{T}^-) + [E(\mathbf{T}^+_{\cdot n}) - E(\mathbf{T}^-)]$$

and in general

$$S_{i,n} = (i-1)E(\mathbf{T}^-) + E(\mathbf{T}^+_{\cdot n}) = iE(\mathbf{T}^-) + [E(\mathbf{T}^+_{\cdot n}) - E(\mathbf{T}^-)]$$

and thus the slope of the serial position curve, $E(\mathbf{T}^-)$, is the same as the slope of the target-absent and present curves.

We now have a fixed-path serial self-terminating model that can predict target-present and target-absent curves that are linear and parallel. Serial position curves are now more reasonable with slopes half their previous value. In return for the more realistic predictions, however, we made the model less intuitively appealing, for it has been forced into the position of having to know display size before the first item begins processing. We already stressed the dubiousness of this assumption.

Note that this fixed-path model cannot predict flat or decreasing serial position curves even if all restrictions are removed from the nontarget rates,

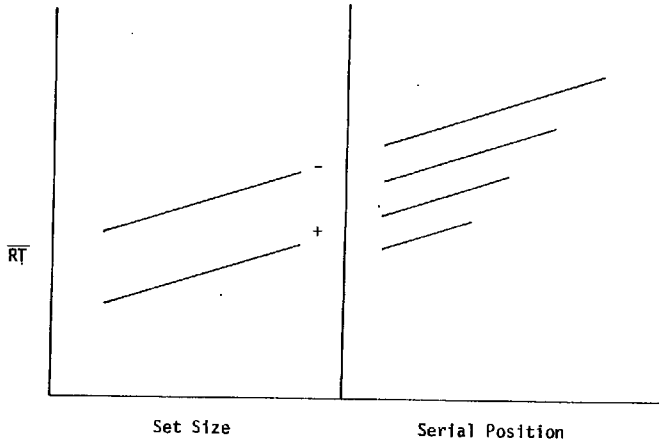


Fig. 7.8. Serial position and set size curves predicted by a serial, self-terminating model with left-to-right search and in which the processing times of target elements are greater the larger the set size.

for flat or decreasing serial position curves imply $S_{1,2} \geqslant S_{2,2}$, which in turn implies $E(\mathbf{T}^+_{\cdot 2}) \geqslant E(\mathbf{T}^-) + E(\mathbf{T}^+_{\cdot 2})$. This restriction can only be fulfilled when nontarget rate is infinite, an alternative we can immediately reject as absurd.

Finally, it can be shown (e.g., Ashby 1976) that the variable-path models, while having no problem with the increasing serial position curves, cannot simultaneously predict parallel target-present and target-absent curves, at least not when only one target and one nontarget rate are assumed. It appears that the extra rate parameters that the fixed path model possesses are essential with this type of data.

Serial exhaustive models

Serial exhaustive models predict that with a single target in the ith serial position of a display containing n elements the mean processing time will be

$$S_{i,n} = E(\mathbf{T}^-_{1,n}) + E(\mathbf{T}^-_{2,n}) + \cdots + E(\mathbf{T}^-_{i-1,n}) + E(\mathbf{T}^+_{i,n}) + \cdots + E(\mathbf{T}^-_{n,n})$$

It is not necessary to consider variable-path exhaustive models. They are mathematically equivalent to serial models that always process in the same order. This is because on every trial all elements are always processed and thus the mean processing time $S_{i,n}$ is just the sum of all the individual element mean processing times and it does not matter in what order this sum is calculated (we continue to assume that the rates do not vary with processing order).

Serial exhaustive models *can* predict serial position effects. If target items

take longer to process the farther to the right they are placed within the display, then serial exhaustive models will predict increasing serial position curves. If targets on the right of the display are processed faster, the serial position curves will decrease. In fact, by an appropriate manipulation of the target rates any serial position curve can be predicted.

When we condition on the serial position of the target and study the behavior of the system under different processing loads, the constraints imposed on exhaustive models are very similar to those that we saw on the self-terminating models. The natural prediction is that processing time will increase as the load is increased. As we previously mentioned, this is also the most commonly observed empirical result. To predict no change or a decrease in processing time with increases in load, exhaustive models, like self-terminating models, must postulate a speedup of processing that depends on the total number of elements to be processed, an assumption that we decided is often counterintuitive.

The predictions of the simplest possible (standard) serial exhaustive model on the averaged target-present and absent curves is well known (e.g., Sternberg 1966). The model, which assumes that all elements require the same average time to be processed, predicts, of course, linear and parallel curves (see Chapter 6). At the same time this model predicts flat serial position curves. If all elements are always processed at the same rate, then processing time cannot depend on where the critical element was placed in the display.

Earlier we mentioned that left-to-right increasing serial position curves are sometimes obtained in ET and LT tasks (e.g., Townsend & Roos 1973). Although the standard serial exhaustive model cannot predict this result, we might naturally ask which serial exhaustive models can. We have already seen that taken singly, neither increasing serial position curves nor linear and parallel target-present and absent curves present a problem for this class of models. As it happens, the combination of these two results can also be predicted (Townsend 1973a; Ashby 1976 presents a greatly expanded treatment). In fact, serial exhaustive models can predict any type of serial position curve (i.e., increasing, decreasing, or flat) and still maintain linear and parallel target-present and target-absent curves. But the cost is very high. A serial exhaustive model that predicts increasing or decreasing serial position curves must postulate at least $n(n+1)/2$ target rate parameters (where n is the maximum display size used) and at least one nontarget rate parameter. There is no way to reduce this number without sacrificing at least one of the predictions (i.e., increasing serial position curves, linear target-present and absent curves, or parallel target-present and absent curves). In a design that employs displays of size 1 through 5, the model must therefore postulate at least 16 rate parameters. A good fit will tell us very little about the underlying processing system. Thus, although this pattern of results does not allow us to rule out a serial exhaustive explanation, whenever we observe either increasing or decreasing serial position curves at the same time there are linear and parallel target-present and absent curves, parsimony suggests we concentrate on testable alternative models.

Independent parallel self-terminating models

Our analysis of independent parallel self-terminating models can be performed quickly and easily since they make the appealing prediction that $S_{i,n}=E(\mathbf{T}^+_{i,n})$. Total processing time is completely determined by the processing time of the target element. Nontarget element processing times have no effect on the serial position predictions. For this reason we can use the serial position curves to quickly estimate the target rates of the model if we wish. We need only set $S_{i,n}$ equal to the mean processing time of the target when it is in serial position i and the load is n elements. Given this mean, we can estimate the parameters of the individual element processing time density functions using a method of moments technique. For instance, with exponential processing times $E(\mathbf{T}^+_{i,n})=1/v^+_{i,n}$, and thus we can estimate $v^+_{i,n}=1/S_{i,n}$.

We see immediately that parallel self-terminating models can predict any set of serial position curves exactly using this estimation procedure. If the serial position curves are predicted exactly, then it must be true that the averaged target-present curve is also predicted exactly. Notice that we have not yet placed any constraints on the nontarget rates. These can therefore be freely adjusted to give an adequate fit to the target-absent curve. Using this estimation procedure, it is clear that we can fit *any* combination of serial position curves and averaged target-present and target-absent curves to any desired degree of accuracy. Of course, in some situations a perfect fit may require an inordinate number of free parameters, but it should be clear that the model is much more flexible on this dimension than are, say, serial exhaustive models.

Independent parallel exhaustive models

If processing is parallel and exhaustive, the mean processing time $S_{i,n}$ is given by

$$S_{i,n}=E[\max(\mathbf{T}^-_{1,n}, \mathbf{T}^-_{2,n}, \ldots, \mathbf{T}^-_{i-1,n}, \mathbf{T}^+_{i,n}, \mathbf{T}^-_{i+1,n}, \ldots, \mathbf{T}^-_{n,n})]$$

Predictions of these models are rather difficult to evaluate. In general, there are many different ways they can generate serial position effects. An example may help to clarify this situation. Suppose we wish to determine under what conditions the model can predict increasing serial position curves, that is, under what conditions $S_{i,n}<S_{i+1,n}$. This condition is equivalent to asking when

$$E[\max(\mathbf{T}^+_{1,n}, \mathbf{T}^-_{2,n}, \ldots, \mathbf{T}^-_{n,n})]<E[\max(\mathbf{T}^-_{1,n}, \mathbf{T}^+_{2,n}, \mathbf{T}^-_{3,n}, \ldots, \mathbf{T}^-_{n,n})]<\cdots$$

If the system is deterministic, that is, if the $\mathbf{T}_{i,n}$ are no longer random variables but are instead fixed quantities, then there is only one set of restrictions that satisfy this inequality. That is, $T^+_{1,n}<T^+_{2,n}<\cdots<T^+_{n,n}$. However, with a stochastic system this is only one possible solution of the inequalities [if we substitute $E(\mathbf{T}^+_{i,n})$ for $T^+_{i,n}$]. For instance, the first inequality holds if and only if

$$\int_0^\infty [1-G^+_{1,n}(\tau)G^-_{2,n}(\tau)\cdots G^-_{n,n}(\tau)]\,d\tau$$

$$<\int_0^\infty [1-G^-_{1,n}(\tau)G^+_{2,n}(\tau)\cdots G^-_{n,n}(\tau)]\,d\tau$$

which can be simplified to

$$\int_0^\infty [G^-_{1,n}(\tau)G^+_{2,n}(\tau)\cdots G^-_{n,n}(\tau)-G^+_{1,n}(\tau)G^-_{2,n}(\tau)\cdots G^-_{n,n}(\tau)]\,d\tau<0$$

or equivalently

$$\int_0^\infty G^-_{3,n}(\tau)G^-_{4,n}(\tau)\cdots G^-_{n,n}(\tau)[G^-_{1,n}(\tau)G^+_{2,n}(\tau)-G^+_{1,n}(\tau)G^-_{2,n}(\tau)]\,d\tau<0$$

and thus a sufficient condition for the inequality to hold is that

$$G^-_{1,n}(\tau)G^+_{2,n}(\tau)<G^+_{1,n}(\tau)G^-_{2,n}(\tau) \quad \text{for all } \tau>0$$

This condition clearly cannot be reduced to a restriction only on means. Variances and other moments could play an important role, too. It thus appears that parallel exhaustive models, although analytically difficult to investigate, may have enough flexibility to predict a substantial diversity of results.

We have now seen that no serial exhaustive model with a manageable number of free parameters can predict monotonically increasing or decreasing serial position curves when target-present and target-absent curves are linear and parallel. This result includes some very large and popular classes of models. In addition to the ones explicitly investigated here are variable-path models, as they are mathematically equivalent to fixed-path models when search is exhaustive. Meanwhile, serial self-terminating models easily predict this data but are forced to postulate target rates that depend on display size, a questionable assumption, especially when a visual search paradigm is utilized. On the other hand, parallel self-terminating models can predict any combination of curves without making unreasonable assumptions about their parameters, whereas parallel exhaustive models also appear able to predict many results.

Overall it appears the combination of linear and parallel target-present and target-absent curves and monotonically increasing serial position curves tends to favor self-terminating over exhaustive models and parallel over serial models, in the sense that, to predict these results, the favored class of models makes fewer and/or more intuitive assumptions about its parameters.

Variances and higher moments

Having now investigated fairly thoroughly the mean RT predictions of some large classes of processing models under differing experimental conditions, we might next naturally ask about the efficiency of higher moments in discriminating between self-terminating and exhaustive search strategies. As the degree of the moment in question increases, so does the estimation problem (e.g., Ratcliff 1979), and thus we must be wary of techniques that rely on cumulants or moments of order 3 or larger. This caution must hold no matter how strong the theoretical development, since the problem is purely statistical in nature.

At any rate, these problems are minimal for second-order moments, and thus we now ask if RT variances can help us discriminate between self-terminating and exhaustive processing. The possibility of nonzero covariance terms greatly increases the complexity of the variance expressions for the different models. Because of this, the types of models we investigate will be of a much simpler nature than in our discussions of mean RT.

Let us start by considering serial models. If search is exhaustive, then all the variance arises from the individual element processing times. Thus, if all intercompletion times are independent and identically distributed with variance var(**T**), then the exhaustive processing time when there are n elements in the display, $\mathbf{T}_n$, has variance $\text{var}(\mathbf{T}_n)=n\cdot\text{var}(\mathbf{T})$ and we see that the processing time variance increases linearly with the number of elements processed.

In self-terminating models, things are a bit more complicated, for now the variance can arise from *two* different sources. For a fixed display size, the number of elements actually processed by the system will vary over trials, due to the self-terminating nature of the search. Thus, whereas variance still arises from the variation in the individual element processing times, it also is increased by the variation in the number of elements actually processed from trial to trial. This leads us to expect that processing time variances for self-terminating models may increase at a faster rate than the variances for exhaustive models as display size is increased. If this is true, then a simple and appealing test of search strategy is easily attained. All we need do is compare the target-present and target-absent RT variance curves from a standard memory-scanning or visual search paradigm. If the target-present curve is steeper than the target-absent curve, a self-terminating search is supported.

These arguments are by no means novel. As early as 1903 J. P. Hylan argued that serial self-terminating variances should be large with respect to parallel variances. In fact, from an examination of RT variances, Hylan concluded that during early trials of visual search type tasks, search is serial and self-terminating, but as the observer becomes more practiced, search becomes more parallel. Hylan seems to have assumed implicitly that parallel processing is exhaustive. Hylan's reasoning was not based on sound and detailed mathematical argument. Nevertheless his work anticipates in spirit the strategy recently proposed by Schneider and Shiffrin (1977).

This is not to say that an exhaustive model can never predict that the target-present variance curve will be steeper than the target-absent variance curve. Intuitive exhaustive models predicting such a crossover can be constructed. For instance, consider an exhaustive model that processes items in a regular and systematic fashion until a target is encountered. At this point

suppose some capacity is diverted away from the comparison process; perhaps it is allocated to response selection. In this case the system has qualities common to both self-terminating and exhaustive models but must still definitely be labeled exhaustive since comparison does not halt with the discovery of a target.

One possible consequence of reducing some of the system's capacity is that the processing times of the remaining elements may become more erratic; that is, the processing time variances may increase (as occurs when the exponential rate decreases). Now when display size is increased and search is exhaustive, on the average more elements will be processed after the target is completed and thus in this model, the processing time variance will be greater than if no target had existed in the display. The prediction follows. The target-present variance curve will increase faster than the target-absent curve as display size is increased.

Before we get too far ahead of ourselves, we should verify our intuitions by actually calculating the self-terminating variance curve. Fortunately, in the case where the intercompletion times are independent and have a common mean and variance the solution is not too hard to obtain. In this case, finding the processing time variance is equivalent to finding the variance of the sum of a random number of independent random processing times $\mathbf{T}_{(1,n)} = \mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_{\mathbf{M}}$, where n is the display size and $\mathbf{M}$ is the actual number of elements processed on a given trial (so that $\mathbf{M} \leqslant n$). The solution is given in the next result.

Proposition 7.7: When there is one target in a display of n elements, the processing time variance of a serial self-terminating model in which all intercompletion times are independent with mean $E(\mathbf{T})$ and variance $\text{var}(\mathbf{T})$ is

$$\text{var}_{ST}[\mathbf{T}_{(1,n)}] = \left(\frac{n+1}{2}\right)\text{var}(\mathbf{T}) + \frac{(n-1)(n+1)}{12}[E(\mathbf{T})]^2$$

Proof: First note that under our assumptions the random variables $\mathbf{M}$ and $\mathbf{T}$ are independent, and since the random times $\mathbf{T}$ all have the same means and variances,

$$E[\mathbf{T}_{(1,n)} \mid \mathbf{M}] = \mathbf{M}E(\mathbf{T}) \quad \text{and} \quad \text{var}[\mathbf{T}_{(1,n)} \mid \mathbf{M}] = \mathbf{M}\,\text{var}(\mathbf{T})$$

Both these expressions are still random because the number of elements completed, $\mathbf{M}$, changes randomly over trials. Taking expected values results in

$$E\{E[\mathbf{T}_{(1,n)} \mid \mathbf{M}]\} = E[\mathbf{M}E(\mathbf{T})] = E(\mathbf{M})E(\mathbf{T})$$

and

$$E\{\text{var}[\mathbf{T}_{(1,n)} \mid \mathbf{M}]\} = E(\mathbf{M})\,\text{var}(\mathbf{T})$$

We now make use of two expressions relating unconditional to conditional expectations (see Parzen 1962: 55):

$$E(\mathbf{Y}) = E[E(\mathbf{Y} \mid \mathbf{X})]$$

and

$$\text{var}(\mathbf{Y}) = E[\text{var}(\mathbf{Y} \mid \mathbf{X})] + \text{var}[E(\mathbf{Y} \mid \mathbf{X})]$$

The first expression says that the mean of $\mathbf{Y}$ is equal to the mean of the conditional mean of $\mathbf{Y}$ given $\mathbf{X}$. The second expression says that the variance of $\mathbf{Y}$ is equal to the mean of the conditional variance plus the variance of the conditional mean.

The first formula immediately gives us the mean of $\mathbf{T}_{(1,n)}$:

$$E[\mathbf{T}_{(1,n)}] = E\{E[\mathbf{T}_{(1,n)} \mid \mathbf{M}]\} = E(\mathbf{M})E(\mathbf{T})$$

We can use the second formula to derive the variance of $\mathbf{T}_{(1,n)}$ by

$$\begin{aligned}\text{var}[\mathbf{T}_{(1,n)}] &= E\{\text{var}[\mathbf{T}_{(1,n)} \mid \mathbf{M}]\} + \text{var}\{E[\mathbf{T}_{(1,n)} \mid \mathbf{M}]\} \\ &= E(\mathbf{M})\,\text{var}(\mathbf{T}) + \text{var}\{E[\mathbf{T}_{(1,n)} \mid \mathbf{M}]\}\end{aligned}$$

We are almost done. We need only calculate

$$\text{var}\{E[\mathbf{T}_{(1,n)} \mid \mathbf{M}]\}$$

First recall that $\text{var}(\mathbf{T}) = E(\mathbf{T}^2) - [E(\mathbf{T})]^2$, so that

$$\begin{aligned}\text{var}\{E[\mathbf{T}_{(1,n)} \mid \mathbf{M}]\} &= E\{[E(\mathbf{T}_{(1,n)} \mid \mathbf{M})]^2\} - \{E[E(\mathbf{T}_{(1,n)} \mid \mathbf{M})]\}^2 \\ &= E\{[\mathbf{M}E(\mathbf{T})]^2\} - [E(\mathbf{M})E(\mathbf{T})]^2 \\ &= E(\mathbf{M}^2)[E(\mathbf{T})]^2 - [E(\mathbf{M})]^2[E(\mathbf{T})]^2 \\ &= \{E(\mathbf{M}^2) - [E(\mathbf{M})]^2\}[E(\mathbf{T})]^2 = \text{var}(\mathbf{M})[E(\mathbf{T})]^2\end{aligned}$$

Therefore, we finally arrive at

$$\text{var}[\mathbf{T}_{(1,n)}] = E(\mathbf{M})\,\text{var}(\mathbf{T}) + \text{var}(\mathbf{M})[E(\mathbf{T})]^2 \tag{7.2}$$

If one target is randomly placed in a display of n elements, then

$$\begin{aligned}P(\text{target processed first}) &= P(\text{target processed second}) \\ &= \cdots = P(\text{target processed last}) \\ &= \frac{1}{n}\end{aligned}$$

regardless of search order. Thus the expected number of elements that must be processed before the target is completed is

$$E(\mathbf{M}) = \sum_{j=1}^{n} j \cdot \frac{1}{n} = \frac{1}{n}\sum_{j=1}^{n} j = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

In a similar fashion we see that

$$E(\mathbf{M}^2) = \sum_{j=1}^{n} j^2 \cdot \frac{1}{n} = \frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6}\right] = \frac{(n+1)(2n+1)}{6}$$

so that

$$\text{var}(\mathbf{M}) = E(\mathbf{M}^2) - [E(\mathbf{M})]^2$$
$$= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{(n-1)(n+1)}{12}$$

Substituting these expressions in Eq. 7.2 yields the result. □

We are now in a position to compare the self-terminating variance with the exhaustive variance to see if the self-terminating variance really does increase at a faster rate. The next result establishes the conditions under which the two curves intersect.

Proposition 7.8: If all intercompletion times are independent with mean $E(\mathbf{T})$ and variance $\text{var}(\mathbf{T})$ and if search is serial, then the exhaustive processing time variance equals the self-terminating processing time variance when

$$n = \frac{6\,\text{var}(\mathbf{T})}{[E(\mathbf{T})]^2} - 1$$

When n is less than this value the exhaustive processing time variance is greater, and when n is larger than this value the self-terminating processing time variance is greater.

Proof:

$$\text{var}_{EX}[\mathbf{T}_{(1,n)}] = \text{var}_{ST}[\mathbf{T}_{(1,n)}]$$

implies that

$$n\,\text{var}(\mathbf{T}) = \left(\frac{n+1}{2}\right)\text{var}(\mathbf{T}) + \frac{(n-1)(n+1)}{12}[E(\mathbf{T})]^2$$

This equality reduces to

$$\text{var}(\mathbf{T}) = \frac{n+1}{6}[E(\mathbf{T})]^2$$

Solving for n yields the main result. The results for larger and smaller n are obtained by replacing the equality with an inequality in the derivation. □

We verified our suspicion. In serial models (at least in some serial models) the self-terminating variance rises more sharply with increases in load than the exhaustive variance. The self-terminating variance is initially less than the exhaustive variance, but as the display size is increased it eventually becomes larger. But note that exactly when this crossover occurs depends on the relationship between the mean and the variance of the individual element processing times, that is, between $\text{var}(\mathbf{T}_i)$ and $E(\mathbf{T}_i)$. If the intercompletion time variance is large relative to the mean, then the self-terminating variance

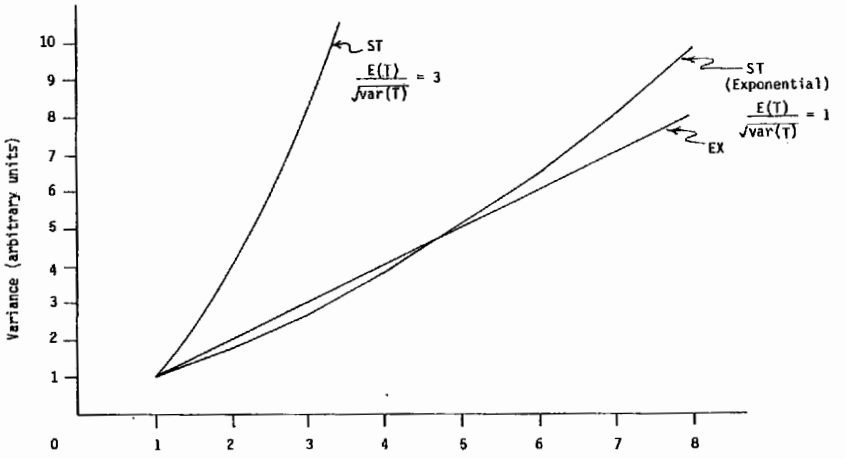


Fig. 7.9. RT variances predicted by serial models as a function of search set size. The self-terminating curve on the right assumes the intercompletion times are distributed exponentially while the curve on the left assumes the intercompletion time variance is smaller (relative to the mean) than in the exponential distribution.

becomes larger than the exhaustive variance only for very large display sizes. Thus, the best discriminability (with regard to this technique) occurs when the individual element processing time variance is small relative to the mean. This is a very plausible result. Remember that the self-terminating variance is generated from two sources: from the variation in the individual element processing times and from the variation in the number of elements processed. Since exhaustive models do not share this second source of variation, the greatest discriminability results from making this second source of variation large relative to the first. This is precisely what happens when the individual element processing time variance is small.

As an illustration of this result, assume the intercompletion times are all exponentially distributed with rate u, so that $E(\mathbf{T}) = 1/u$ and $\text{var}(\mathbf{T}) = 1/u^2$. In this case, the two variance curves cross when

$$n = \frac{6(1/u^2)}{1/u^2} - 1 = 5$$

The complete predictions of the models are given in Fig. 7.9. In typical memory-scanning and visual search tasks, display sizes are rarely larger than 5 or 6 elements, and over this range the curves are very close together, almost certainly too close for empirical discriminations to be possible.

Unfortunately, for very large display sizes, where the variances offer promise as discriminators of search strategy, the question of whether processing is self-terminating or exhaustive becomes trivial. Clearly when lists

become long enough so that reading behaviors begin to dominate, search is more likely to be self-terminating.

On the other hand, the variance of an exponential distribution is fairly large relative to its mean and so with many distributions discriminability will be significantly better than in this example. For instance, for the sake of illustration assume that all increases in the RT mean and variance caused by adding an item to the memory set in a standard memory-scanning experiment (see, e.g., Chapter 6) occur because an extra intercompletion time has been added to the comparison time. This is equivalent to assuming that increasing the memory set from n to $n+1$ items has no effect on any base time process (such as encoding and response selection) and that it does not affect the comparison time of the original n items. If we also assume that all intercompletion times are independent and identically distributed, then we can use the slope of the memory set size vs. mean RT curve (on "no" trials) as an estimate of the mean intercompletion time $E(\mathbf{T})$, and the slope of the memory set size vs. RT variance curve as an estimate of $\text{var}(\mathbf{T})$.

Sternberg (1964) obtained estimates of the first four moments of **T** using just this technique, and in contrast to our exponential example, in which the mean equals the standard deviation, he estimated the mean of **T** to be about three times larger than its standard deviation. The exact predictions of a serial self-terminating model with these parameter values are also given in Fig. 7.9. As predicted by Proposition 7.8, the self-terminating and exhaustive variances diverge for all $n>1$ and thus discriminability between these two strategies is fairly good.

An examination of RT variances can be very helpful in deciding between self-terminating and exhaustive search strategies, but as we have seen, it is not foolproof. A target-present RT variance curve that increases with memory set size at a significantly faster rate than the corresponding target-absent curve is good evidence for a self-terminating search. On the other hand, if the two curves increase at about the same rate, no firm conclusions can be drawn without some knowledge about the ratio of the intercompletion time mean to its variance.

Before concluding this discussion we briefly try to imagine how redundant targets in the display might affect the variance predictions. First of all, if targets and nontargets have the same processing time variance [i.e., if $\text{var}(\mathbf{T}^+)=\text{var}(\mathbf{T}^-)=\text{var}(\mathbf{T})$], then clearly the exhaustive predictions are not affected by the number of targets in the display. Self-terminating predictions will depend on this number and in the manner specified by the next result.

Proposition 7.9: When there are k targets in a display of n elements, the processing time variance of a serial self-terminating model in which all intercompletion times are independent with mean $E(\mathbf{T})$ and variance $\text{var}(\mathbf{T})$ is

$$\text{var}_{ST}[\mathbf{T}_{(k,n)}]=\left(\frac{n+1}{k+1}\right)\text{var}(\mathbf{T})+\frac{k(n-k)(n+1)}{(k+1)^2(k+2)}[E(\mathbf{T})]^2$$

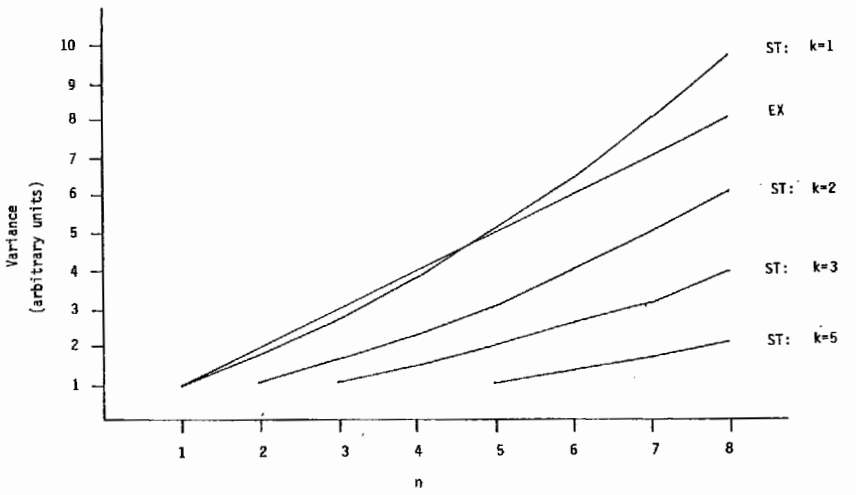


Fig. 7.10. RT variances predicted by serial models as a function of search set size (n) and the number of targets in the search set (k). The intercompletion time distribution is assumed to be exponential.

Proof: First note that Eq. 7.2 is still valid. In the present case, however, the random number of items completed on a single trial, **M**, will have a different mean and variance. It turns out that these expectations are (e.g., Patil and Joshi 1968)

$$E(\mathbf{M})=\frac{n+1}{k+1} \quad \text{and} \quad \text{var}(\mathbf{M})=\frac{k(n-k)(n+1)}{(k+1)^2(k+2)}$$

Substituting these expressions into Eq. 7.2 yields the result. □

To get some idea as to whether redundant targets increase or decrease exhaustive-self-terminating identifiability, let us again assume that all intercompletion times are exponentially distributed with rate u. The predictions of the two classes of models are given in Fig. 7.10 for the case when k, the number of redundant targets, ranges from 1 to 5.

The general effect of redundant targets is seen to be a flattening of the self-terminating curve, and this will be true, not only for the exponential, but for any intercompletion time distribution. The self-terminating curve still always overtakes the exhaustive curve, but as k increases, the display size for which the curves cross will greatly increase. For instance, with 5 redundant targets the two curves cross when n is about 46. Thus, self-terminating and exhaustive predictions are even harder to discriminate when redundant targets are present at the variance level.

We shall finish our discussion of variances by briefly considering the predictions of independent parallel models. If search is parallel and self-

terminating and one target is in the display, then it is the only element that affects processing time. Nontargets do not directly affect RT. Thus adding nontargets to a display containing a target element does not directly affect the RT mean or variance. Of course, increasing the processing load by adding nontargets to the display may alter the amount of capacity allocated to the target element and thereby cause a change in the processing time variance. It is clear, however, that parallel self-terminating models, just as they can predict any relationship among the means, can (for the same reason) predict any relationship between processing load and RT variance.

Again, parallel exhaustive model predictions are more complex. Now we need to calculate the variance of the maximum of a set of random processing times. In general, this is not an easy calculation. Presently we will consider only the case where the individual element processing times have identical exponential distributions. Assume all n elements are processed independently with rate v. Then, as we saw in Chapters 4 and 6, all intercompletion times are independent and exponentially distributed. The first intercompletion time has rate nv, the second has rate $(n-1)v$, and the jth intercompletion time has rate $(n-j+1)v$. Because the intercompletion times are independent and additive, we know the processing time variance is the sum of the intercompletion time variances. Thus[3]

$$\text{var}_{EX}[\mathbf{T}_{(1,n)}] = \left[\frac{1}{nv}\right]^2 + \left[\frac{1}{(n-1)v}\right]^2 + \cdots + \left[\frac{1}{v}\right]^2$$

$$= \frac{1}{v^2}\sum_{j=1}^{n}\frac{1}{j^2}$$

This function is illustrated in Fig. 7.11, where the variance predictions of other unlimited capacity (at the individual element level) exponential models are also displayed. It can be seen that it increases very slowly with n. In fact, since $\sum_{j=1}^{n}(1/j^2)$ converges as n approaches infinity, we see that $\text{var}_{EX}[\mathbf{T}_{(1,n)}]$ is a bounded function for parallel exhaustive models of this type. Note that the parallel self-terminating function is flat, reflecting the capacity structure of the system. Since processing is assumed to be of unlimited capacity, the target element is always processed with the same rate regardless of display size. It follows that the processing time variance (determined only by target element processing times) is invariant with respect to changes in n. Of course, limited capacity parallel models can mimic the serial prediction.

Only a very few studies have reported RT variance vs. set size curves for both target-present and target-absent conditions. Unfortunately, the results appear to be equivocal. For example, in a study employing simultaneous

[3] The variance can also be obtained from the moment-generating function, which can be shown to be equal to

$$M_n(\theta) = \sum_{j=1}^{n}(-1)^{j+1}\frac{(nv)\binom{n-1}{j-1}}{jv+\theta}$$

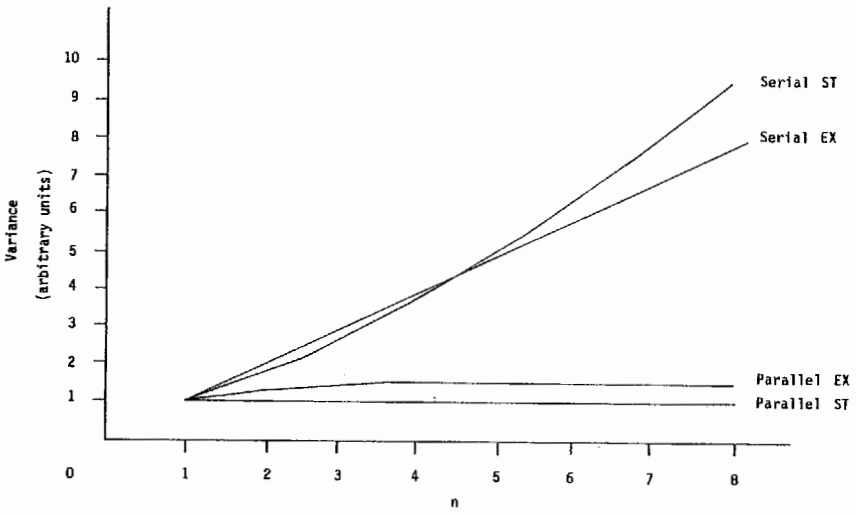


Fig. 7.11. RT variances predicted by various unlimited capacity models with exponentially distributed intercompletion times.

memory and visual search (see Chapter 6) Schneider and Shiffrin (1977) found a slight tendency for the target-present RT variances to increase faster than the target-absent variances as the total number of comparisons an exhaustive model would perform (i.e., the number of items in the memory set times the number in the visual set) increased. Based on this finding, they argued for a serial self-terminating search (or an equivalent parallel process). Using the same experimental paradigm, Rossmeissl et al. (1979) found the exact opposite result, namely a slight tendency for the target-absent curve to increase faster than the target-present curve, and on this basis they favored a parallel explanation. The theoretical results of this section certainly seem encouraging enough to warrant further empirical investigation.

Distributional approaches

There have been a few attempts to devise methods that attempt to discriminate between self-terminating and exhaustive search strategies by utilizing the entire RT distributions. These methods were developed by first considering a certain class of models delineated by some very specific assumptions and then by deriving from these assumptions certain properties or testable predictions about the RT distributions.

Sternberg (1973) proceeded in this fashion and proposed two properties of the RT cumulative distribution functions that he showed were predicted by a large class of self-terminating models. Sternberg's approach is important, and because of the potential generality of the class of models he considers, we examine it in detail in the next chapter (see also Vorberg 1981).

A somewhat different approach was taken by Ashby and Townsend (1980). They assumed a model of memory-scanning or visual search in which the following assumptions hold:

(A1) The base time is independent of the comparison time.

(A2) Within the comparison time all intercompletion times are independent and exponentially distributed.

(A3) The intercompletion time rates do not depend on display size.

Note that this model does not require all intercompletion times to be identically distributed. The intercompletion times may thus all unfold at different rates.

Let us denote the observable RT density when n elements are in the display by $f_n(t)$ and the associated cumulative distribution function by $F_n(t)$. Finally let the rate of the nth intercompletion time be V_n (i.e., the nth intercompletion time density is $V_n e^{-V_n t}$). Then Ashby and Townsend showed (see Proposition 3.8) that if search is exhaustive, the nth intercompletion time rate can be recovered by

$$V_n = \frac{f_n(t)}{F_{n-1}(t) - F_n(t)} \quad \text{for any } t > 0 \tag{7.3}$$

Note that this relation holds for every time t. Thus, no matter what value of t is chosen, the same estimate of V_n is attained. On target-absent trials an exhaustive search is guaranteed, and so a plot of the right-hand side of Eq. 7.3 vs. t will test the other assumptions of the model. If the resulting plot is linear and independent of time, then the assumptions are supported and a test between exhaustive and self-terminating search strategies can be conducted.

On target-present trials and if search is self-terminating, Eq. 7.3 will no longer hold. This is because with n elements in the display there are no longer n intercompletion times. In fact, Ashby and Townsend (1980) showed that if assumptions (A1) through (A3) hold and if search is self-terminating and the right side of Eq. 7.3 is plotted against time, then the result will not be a linear function independent of time.

This result allows us to test for search strategy. If we plot Eq. 7.3 vs. t for the target-absent data and the result is a flat linear function (i.e., zero slope), then we know the three assumptions of the model are supported. If we repeat this procedure for the target-present data with the same result, then we know that search cannot be self-terminating and that an exhaustive search is supported. If, instead, the resulting plot is not linear and flat, then we conclude processing is not exhaustive. This conclusion is reached because we assume that if assumptions (A1) through (A3) hold for the target-absent data, they will also hold for the target-present data. Changes in the slope and linearity of the plot are therefore attributed to changes in search strategy (e.g., from exhaustive to self-terminating). In such a case, self-termination may be suspected, but what we would like is a way to verify this intuition. Ashby and Townsend (1980) provided such an opportunity, for they showed that if

search is self-terminating, then the nth intercompletion time can still be recovered, but this time by

$$V_n = \frac{nf_n(t) - (n-1)f_{n-1}(t)}{2(n-1)F_{n-1}(t) - (n-2)F_{n-2}(t) - nF_n(t)} \quad \text{for any } t > 0 \tag{7.4}$$

It is reiterated that Eqs. 7.3 and 7.4 are useful not only as a means of estimating the intercompletion time rates when assumptions (A1) through (A3) hold, but also as a test of this set of assumptions. Using Eqs. 7.3 and 7.4 to discriminate between self-terminating and exhaustive search strategies is possible only if the entire set of three assumptions is supported.

Testing paradigms based on the entire RT distributions appear extremely promising. They utilize all of the RT data, make strong, clear-cut predictions, and generally provide a means of at least partially testing their assumptions. It is the severity of the assumptions these methods postulate, however, that prevents one from heralding a new era in RT theorizing. Given enough assumptions, very specific and powerful predictions can always be made. The promising future of these approaches foretells a struggle to relax assumptions and generalize the models and yet still retain the powerful and specific predictions.

Tests involving accuracy

We have examined many ways that purport to discriminate between self-terminating and exhaustive search strategies. Overall the emerging picture looks favorable, suggesting that this processing dimension may offer more immediate observability (and testability) of the system than does, say, the parallel–serial dimension. The paradigms based on mean RT appear the more powerful of those we have discussed, possibly because they have received a great deal more attention in the literature.

The major criticism of the results of this chapter is probably that we have considered only those instances in which accuracy losses are inconsequential, since all of the methods so far discussed assume that all relevant information about the search strategy is contained in the RT data and thus that the accuracy data can be ignored without serious consequences. We believe this is a valid strategy in experimental situations in which accuracy is very high or exhibits no interesting covariance with RT. However, when accuracy losses are substantial, and especially when these losses are correlated in some way with RT, then the assumption becomes suspect.

A self-terminating search strategy is usually thought to be more efficient than an exhaustive strategy with regard to the system's total energy consumption, because on the average fewer elements need to be processed. On error-free trials, this energy savings manifests itself in the overall shorter RTs that will be observed. If performance is not perfect, the more efficient self-terminating search strategy may result in an increase in accuracy and/or a

decrease in RT relative to less efficient search strategies. If we were to ignore accuracy data and instead concentrate on latencies, and if the energy saved by self-termination resulted primarily in accuracy gains, then one might mistakenly conclude that search was exhaustive.

What we need is some observable measure of the average energy consumed by the system over trials. We could then compare the values obtained on the target-present and the target-absent trials as display size is increased. If the rate of increase in energy consumed is about the same on both types of trials, then an exhaustive search is supported, whereas if proportionately less energy is consumed on target-present trials, a self-terminating search is supported. We want this measure to increase with processing time and to decrease with improvements in accuracy, and we want to be able to compute it for each experimental condition. As a measure monotonically related to processing time, we can use mean RT. A desirable accuracy measure is often a statistic such as d', the estimate of the detectability parameter from standard signal detection theory. Unfortunately, this parameter cannot be estimated in any single experimental condition. For example, in the target-absent conditions we can estimate the probability of a false alarm but not of a hit. An estimate of d' requires both these probabilities. Thus we must make the best of what we have. As a compromise, we might use the proportion of correct responses, $P(\text{corr})$, as our accuracy measure.

Next we must decide what function of these two statistics we shall take as our estimate of the consumed energy. Townsend and Ashby (1978) suggest the ratio of the two, and for illustrative purposes we adopt that notion here. Figure 7.12a shows a plot of $\overline{\text{RT}}/P(\text{corr})$ vs. display size that supports an exhaustive search, and Fig. 7.12b supports a self-terminating search.

Note that the curves look very much like the mean RT vs. display size curves predicted for standard serial models in an ET or LT task. The main effect of a positive correlation between errors and display size will be an increase in slope. The linearity of the curves will not be affected so long as the accuracy decrements are linearly related to display size.

Although a consideration of the statistic $\overline{\text{RT}}/P(\text{corr})$ will yield more information than just examining mean RT, the best measure is one tied to the structure of the assumed processing model in such a way that an inverse transformation carries one back from the speed–accuracy curve to the curve relating the instantaneous energy (i.e., power) consumed by the system to time. Assuming a constant power consumption of the system over time allows us to examine the structure of the system. If the inverse transformation does not indicate a constant power output, then we have assumed the wrong structure for the system. Townsend and Ashby (1978) discuss this issue at length.

The next chapter, which deals in detail with distribution orderings, emphasizes the capacity issue. In it, we shall have occasion to present an approach to the study of capacity based on certain nonparametric properties of the processing time distributions. Interestingly, this development has its origins in the self-terminating vs. exhaustive controversy, since the most important of the properties we will discuss were originally intended as tests of the self-terminating hypothesis. We return to accuracy in relation to RT in Chapter 9.

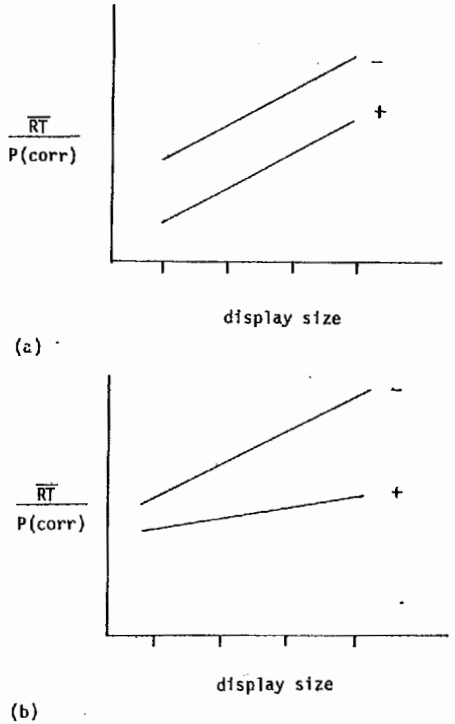
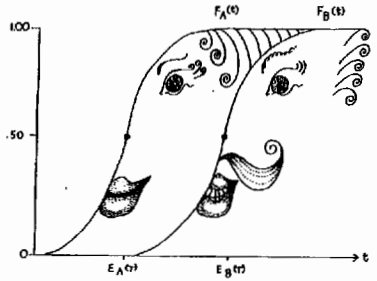


Fig. 7.12. $\overline{\text{RT}}/P(\text{correct})$ as a function of search set size for target-present (+) and target-absent (−) conditions. (a) shows results representative of exhaustive models and (b) shows results representive of self-terminating models.



A graph illustrating the concept of *probability distribution function ordering*. Time is on the x axis or abscissa, while probability is on the y axis or ordinate. Two cumulative probability distribution functions, $F_A(t)$ and $F_B(t)$, are exhibited here. Note that always $F_A(t) \geq F_B(t)$, thus revealing a distribution ordering. $E_A(t)$ and $E_B(t)$ are the means or expectations of the respective distributions (in this instance they equal the medians as seen by the position of the .50 points). The faces are for decoration.

8 *Nonparametric RT predictions: distribution-ordering approaches*

Introduction

In this chapter we consider some approaches to the study of processing systems that are based on nonparametric properties of the processing time distributions. As such, these approaches are quite broad in that they are potentially relevant to large classes of models. Although their development shall necessitate some assumptions about the processing system, it will not be necessary to assume a specific distribution (e.g., exponential) on, say, the intercompletion times or the actual processing times.

We shall attempt to relate these nonparametric properties directly to certain fundamental principles of processing; in particular, the concept of capacity will play an especially integral role in the developments that follow. We can then test a set of data for the presence of these properties and thus for the existence of the associated processing principle. This in turn will enable us to delimit, for that set of data, the class of potential explanatory models.

In this first section of the chapter, the basic concept of an order placed on a set of distribution functions is introduced. Throughout the bulk of the chapter we will use these orderings to make inferences about the capacity of the system. In the next two sections, the orderings will be used to study

capacity at the exhaustive and self-terminating levels of processing. In the fourth section, capacity at the individual element level is treated and definitions are given that relate exhaustive, self-terminating, and minimum time processing to capacity at the individual element level.

The fifth section presents a test of self-termination proposed in an address by Sternberg (1973). The primary reason why the test is presented here and not in the preceding chapter is that the two test properties Sternberg recommends, the long- and short-RT properties, are both in the form of an ordering on a set of distribution functions as the processing load is changed, and so we believe they are intimately related to the capacity structure of the system as well as to whether search is self-terminating or exhaustive.

This section includes a major investigation of the class of independent parallel models with regard to the two test properties. Specifically, two questions will be considered in detail. First, given knowledge only of the capacity level of a system, is it possible to predict whether one or both of the RT properties will be violated? Secondly, and more importantly to the empiricist, given some specific outcome to an empirical test of the two properties (e.g., long-RT property holds but short-RT property fails), what can we infer about the processing system's capacity structure? It will be shown that much indeed can be inferred, and thus the long- and short-RT properties are potentially powerful indicants of processing capacity. The preceding developments are then applied to a set of data gathered by Townsend and Roos (1973).

Finally, the last section of the chapter investigates capacity during the first stage or intercompletion time of processing. We will show that instead of using orderings on distribution functions, as we do in the rest of the chapter, at this level of processing more compelling results are obtained if capacity is defined through an ordering on sets of hazard functions (see also Chapter 3).

Throughout this chapter, we assume that accuracy is controlled in such a way that RT alone can be used to draw conclusions about capacity and other processing characteristics. For instance, we may compare the sets of processing times for two different display sizes in a memory-scanning task, given that the accuracy levels are about the same and relatively small for both sets.

To begin, let us assume that we have available the processing time distributions for some level of processing under two different processing loads. Let $f_A(t)$ be the probability density function of times for whatever level of processing we have decided to observe when there are few elements to be processed, and let $f_B(t)$ be the density of times for that same level but when more elements have been added to the load. Thus $f_A(t)$ might be, for example, the first intercompletion time density function when there are only one or two elements to be processed, and $f_B(t)$ the first intercompletion time density when there are five or six elements. Note that these densities are not necessarily serial in the present context (until future notice). The *level* of processing here is the first-stage duration.

The problem now facing the investigator is to decide, on the basis of $f_A(t)$ and $f_B(t)$, the capacity of the system. Thus for example, if $f_A(t)$ and $f_B(t)$

are identical, we know that increasing the load did not, in this case, affect the processing level we observed, and so we must conclude that the system is unlimited capacity *at that level* (e.g., first intercompletion time level). If the densities are not identical, the problem is often not so simple. One can, of course, just compute the respective means and compare these, but this ignores much valuable information contained in the density functions.

Townsend and Ashby (1978) developed measures that offer other solutions to this problem besides just comparing the means. One alternative is to compare the probabilities, under the different loads, that processing has already been completed by some time t. This is formally equivalent to comparing cumulative distribution functions and looking for orderings of the kind $F_A(t) > F_B(t)$ for all $t > 0$. A second possibility is to examine an ordering based on the random times themselves. That is, one could ask if the probability that the random time $\mathbf{T}_A$ is less than the random time $\mathbf{T}_B$ is itself greater than one-half; i.e., is $P(\mathbf{T}_A < \mathbf{T}_B) > \frac{1}{2}$?

As it turns out, a distribution function ordering is stronger than an ordering based on the random times, in the sense that $F_A(t) > F_B(t)$ for all $t > 0$ implies $P(\mathbf{T}_A < \mathbf{T}_B) > \frac{1}{2}$ but that the reverse implication does not hold. Further, an ordering on the distribution functions implies an ordering on means and medians, whereas $P(\mathbf{T}_A < \mathbf{T}_B) > \frac{1}{2}$ does not imply an ordering on means or medians or vice versa. Thus, means, medians, and $P(\mathbf{T}_A < \mathbf{T}_B) > \frac{1}{2}$ are at about the same predictive strength.

Capacity in exhaustive processing

In this sense, then, we can see that the distribution functions are the more powerful measures of capacity.[1] Again, though, before one can implement a study of capacity via, say, distribution functions, the level of processing to be studied must be selected. If, for example, we are interested in exhaustive processing, then $f_A(t)$ and $f_B(t)$ should be overall processing time densities for those trials on which we expect search to be exhaustive.

If $f_A(t)$ is the processing time density when the display contains one element, and $f_B(t)$ when two elements are processed, we conclude that exhaustive processing is limited capacity on the distribution functions if $F_A(t) > F_B(t)$ for all $t > 0$, that it is supercapacity if $F_A(t) < F_B(t)$, and otherwise that it is unlimited capacity [i.e., if $F_A(t) = F_B(t)$]. Suppose now that on some application of this procedure we find evidence of limited capacity at the exhaustive level. We might now ask what this tells us about the behavior of the system at the level of the intercompletion time.

[1] There are, as it turns out, orderings even stronger than those on the distribution functions. For instance, Townsend and Ashby (1978) showed that an ordering on hazard functions orders the distribution functions but not vice versa. In Chapter 9 (see Proposition 9.5) we prove that monotonic likelihood ratios imply an ordering on the hazard functions and are thus stronger yet. As we indicated, the possibility of using hazard functions as measures of capacity will be taken up later in this chapter.

To simplify the discussion we restrict ourselves to models that assume independence of the individual element processing times, and we will continue using the notation of the preceding chapter. Thus $\mathbf{T}_{i,n}$ will refer to the random time it takes a serial system to process the element in the ith position in a display of n elements, and $\mathbf{T}_{i,n}$ is the corresponding time if processing is parallel. As before, the serial density function corresponding to this random processing time is $f_{i,n}(t)$, whereas in a parallel system it is $g_{i,n}(t)$. To facilitate comparison between serial and parallel models, we will use the symbol t rather than τ throughout this chapter to refer to actual parallel processing times.

We begin by investigating serial models. When search is exhaustive the overall density function can be written, using our convolution notation, as

$$f_{\mathrm{EX}_n}(t) = f_{1,n}(t) * f_{2,n}(t) * \cdots * f_{n,n}(t)$$

From this expression, the distribution function can be found to be

$$\begin{aligned} F_{\mathrm{EX}_n}(t) &= \int_0^t f_{\mathrm{EX}_n}(t')\,dt' \\ &= F_{1,n}(t) * f_{2,n}(t) * f_{3,n}(t) * \cdots * f_{n,n}(t) \end{aligned}$$

If the individual element processing times (which for serial models are the same as intercompletion times) are exponentially distributed, then the serial exhaustive processing time has the general gamma, or the Erlangian distribution (see, e.g., Chapter 3; McGill & Gibbon 1965), whose density is

$$f_{\mathrm{EX}_n}(t) = \sum_{i=1}^{n} C_{i,n} u_{i,n} \exp(-u_{i,n} t) \quad \text{for } t \geq 0$$

Here $u_{i,n}$ represents the rate of the ith of the n stages. The constant $C_{i,n}$ is defined as

$$C_{i,n} = \prod_{\substack{j=1 \\ j \neq i}}^{n} \frac{u_{j,n}}{u_{j,n} - u_{i,n}}$$

Since the density function is just a weighted sum of exponentials, the integration needed to find the distribution function is easily performed, yielding

$$F_{\mathrm{EX}_n}(t) = 1 - \sum_{i=1}^{n} C_{i,n} \exp(-u_{i,n} t)$$

These expressions are true only if all elements have different processing rates, that is, only if $u_{i,n} \neq u_{j,n}$ for all $i \neq j$. In the case where some of the rates are equal, another equation for the RT density and distribution function must be used. For example, when two elements both have the rate u,

$$F_{\mathrm{EX}_2}(t) = 1 - e^{-ut}(ut + 1)$$

Next, let us consider parallel exhaustive processing. As stated, we will restrict the discussion to models that assume independent total completion

times (see Chapters 3 and 4). Under this assumption the joint probability that a is finished at time $\mathbf{T}_a \leqslant t_a$ and b at time $\mathbf{T}_b \leqslant t_b$ can be written as

$$P(\mathbf{T}_a \leqslant t_a, \mathbf{T}_b \leqslant t_b) = P_a(\mathbf{T}_a \leqslant t_a) P_b(\mathbf{T}_b \leqslant t_b)$$
$$= G_a(t_a) G_b(t_b)$$

With n elements to be processed, the time to complete them all will be the maximum of the n individual processing times for the various serial positions. Thus the probability that the exhaustive processing time is less than t, $G_{\mathrm{EX}_n}(t)$, is equal to the probability that all individual item processing times are less than t, and so

$$G_{\mathrm{EX}_n}(t) = G_{1,n}(t) G_{2,n}(t) \cdots G_{n,n}(t) = \prod_{i=1}^{n} G_{i,n}(t)$$

When the individual element processing times are all exponentially distributed,

$$G_{\mathrm{EX}_n}(t) = \prod_{i=1}^{n} [1 - \exp(-v_{i,n} t)]$$

It might be interesting to compare the serial and parallel distribution functions when one element is being processed to the exhaustive distribution function when two elements are processed in the case when all elements are processed with the same rate. In this case we have unlimited capacity at the *individual element level* because the individual element processing rates are the same under both loads. The following result shows that for both serial and parallel processing, this situation leads to limited capacity at the exhaustive processing level.

Proposition 8.1: With unlimited capacity at the individual element level,

$$F_{\mathrm{EX}_1}(t) \geqslant F_{\mathrm{EX}_2}(t) \quad \text{for all } t > 0$$

and

$$G_{\mathrm{EX}_1}(t) \geqslant G_{\mathrm{EX}_2}(t) \quad \text{for all } t > 0$$

Proof: We interpret unlimited capacity at the individual element level to mean that the processing time of each element has density function $f(t)$ in the case of serial processing and $g(t)$ in the case of parallel processing.

(Serial proof) According to this definition,

$$F_{\mathrm{EX}_2}(t) = F(t) * f(t) = \int_0^t F(t-t') f(t')\, dt'$$

$$\leqslant F(t) \int_0^t f(t')\, dt' \leqslant F(t) = F_{\mathrm{EX}_1}(t)$$

and the inequalities must hold for all $t > 0$.

(Parallel proof) In the case of parallel processing,

$$G_{\mathrm{EX}_2}(t) = [G(t)]^2$$
$$\leqslant G(t) = G_{\mathrm{EX}_1}(t) \quad \text{for all } t > 0 \ \square$$

This result illustrates that our conclusions about the capacity of a system will often depend upon which level of processing we are investigating. In the above result, we found limited capacity for the distribution functions at the exhaustive level, even when the individual element processing rates were unaffected by an increase in the load, that is, even though the individual element capacity was unlimited.

The relationships among the different levels will depend on the characteristics of the system under scrutiny. For instance, unlimited capacity of a deterministic parallel system at the level of the individual element implies unlimited capacity at the level of exhaustive processing. However, no such implication holds for probabilistic parallel systems, as we learned in Chapter 4. Therefore, it is critical to specify the processing level of interest and often other aspects of the system as well. We shall proceed with this in mind.

Before moving on to self-terminating stopping rules, it is of interest to compare the serial and parallel exhaustive distribution functions to get some idea as to which processing strategy is more efficient. Proposition 8.2 specifically considers the case where the individual element processing times are identically distributed in the two models.

Proposition 8.2: If the individual element processing times in serial and parallel models are identically distributed, then $G_{\mathrm{EX}_2}(t) \geqslant F_{\mathrm{EX}_2}(t)$ for all $t > 0$.

Proof: Suppose the individual item processing time density function in the two models is $k(t)$. Then

$$F_{\mathrm{EX}_2}(t) = K(t) * k(t) = \int_0^t K(t-t') k(t')\, dt'$$

$$\leqslant K(t) \int_0^t k(t')\, dt' = [K(t)]^2 = G_{\mathrm{EX}_2}(t)$$

and the inequality must hold for all $t > 0$. $\square$

This result says that for a given individual element processing rate, parallel exhaustive processing is, on the whole, faster than serial exhaustive processing, given equal independent rates on the elements being processed. This is as we intuited: Doing things simultaneously is faster than doing things sequentially. A rather different question, however, is whether a parallel system can ever process many items, each at the same rate that a serial system processes only one item. We shall consider this question in some detail when

we study capacity at the level of the first stage of processing later in this chapter. However, our attention will now turn to self-terminating stopping rules.

Capacity in self-terminating processing

In some experimental situations there are trials on which a single target or critical element containing all relevant information about that trial is embedded in the stimulus array. It may be that on these trials the system terminates processing as soon as the critical element is discovered, regardless of how many elements remain uncompleted. In such cases we say that search is self-terminating (see Chapters 4 and 7). Since, on such trials, the observer does not typically process all elements in the display, our exhaustive distribution functions will not describe the system's behavior in these circumstances. Fortunately, the appropriate distribution functions are easily written.

Assume processing is serial and that position one is always processed first. If the target is randomly placed within the display, then no matter what the search order, all possible *completion* orders will be equally likely. Thus, for two elements (recalling that $f_{i,j}(t)$ is the density for position i when j elements are in the display and letting $f_{i+k,j}(t)$ equal the density when both position i and k must be processed),

$$f_{ST_2}(t)=\tfrac{1}{2}[f_{1,2}(t)+f_{1+2,2}(t)]$$

By our across-stage independence assumption this expression reduces to

$$f_{ST_2}(t)=\tfrac{1}{2}\{f_{1,2}(t)+[f_{1,2}(t)*f_{2,2}(t)]\}$$

Thus, the distribution function can be written as

$$\begin{aligned}F_{ST_2}(t)&=\tfrac{1}{2}[F_{1,2}(t)+F_{1+2,2}(t)]\\&=\tfrac{1}{2}\{F_{1,2}(t)+[F_{1,2}(t)*f_{2,2}(t)]\}\end{aligned}$$

If the system does not always search the display by beginning in position 1, the expression must be modified somewhat. Assume position 1 is searched first with probability p. Then

$$\begin{aligned}f_{ST_2}(t)=&\tfrac{1}{2}\{pf_{1,2}(t)+(1-p)[f_{1,2}(t)*f_{2,2}(t)]\}\\&+\tfrac{1}{2}\{(1-p)f_{2,2}(t)+p[f_{1,2}(t)*f_{2,2}(t)]\}\end{aligned}$$

The first term applies to trials on which the target is in position 1, while the second term describes search time on trials when the target is in position 2. The above expression can be simplified to

$$f_{ST_2}(t)=\tfrac{1}{2}[pf_{1,2}(t)+(1-p)f_{2,2}(t)]+\tfrac{1}{2}[f_{1,2}(t)*f_{2,2}(t)]$$

and thus the distribution function can be written as

$$F_{ST_2}(t)=\tfrac{1}{2}[pF_{1,2}(t)+(1-p)F_{2,2}(t)]+\tfrac{1}{2}[F_{1,2}(t)*f_{2,2}(t)]$$

In the special exponential case, when the processing rate does not depend on serial position,

$$F_{ST_2}(t)=\tfrac{1}{2}[1-\exp(-u_2t)]+\tfrac{1}{2}[1-\exp(-u_2t)(1+u_2t)]$$

When processing is parallel and independent, things are especially simple. The self-terminating processing time is the same as the target element processing time. Nontarget element processing times are irrelevant (although the number of nontarget elements may affect the target rate). To calculate the overall self-terminating distribution function $G_{ST_n}(t)$, we assume that the target occurs equally often in all positions. Then $G_{ST_n}(t)$ is simply the mean of the target element processing time distribution functions across the different serial positions,

$$G_{ST_n}(t)=\frac{1}{n}\sum_{i=1}^{n}G_{i,n}(t)$$

If the target distributions are exponential, then

$$G_{ST_n}(t)=\frac{1}{n}\sum_{i=1}^{n}[1-\exp(-v_{i,n}t)]$$

As we did with exhaustive processing, we can compare the self-terminating distribution functions for serial and parallel models across different processing loads when capacity is unlimited at the individual element level.

Proposition 8.3: With unlimited capacity at the individual element level, $F_{ST_1}(t)\geqslant F_{ST_2}(t)$ for all $t>0$, but $G_{ST_1}(t)=G_{ST_2}(t)$ for all $t>0$.

Proof: Our interpretation of unlimited capacity at the individual element level is the same as in Proposition 8.1.

(Serial proof) With unlimited capacity at the individual element level, $F_{ST_1}(t)=F(t)$, whereas $F_{ST_2}(t)=\tfrac{1}{2}F(t)+\tfrac{1}{2}[F(t)*f(t)]$. Therefore, the result holds if $F(t)\geqslant F(t)*f(t)$ for all $t>0$. This ordering was verified in the proof of Proposition 8.1.

(Parallel proof) With unlimited capacity at the individual element level,

$$\begin{aligned}G_{ST_2}(t)&=\tfrac{1}{2}[G(t)+G(t)]\\&=G(t)=G_{ST_1}(t)\quad\square\end{aligned}$$

The serial result is the same as we found with an exhaustive search: When capacity is unlimited at the individual element level, it is limited at the self-terminating level. However, in the case of parallel processing, unlimited capacity at the individual element level implies *unlimited* capacity at the level of self-termination, something we have not yet seen. On the other hand, this result might have been expected since no matter how many elements are in the display, the total processing time is equal to the completion time of the one

pertinent target element. If the target-processing time is independent of the number of nontargets in the display (as in unlimited capacity at the individual element level) we expect self-terminating processing to be independent of load.

Another comparison we made when discussing exhaustive search was between parallel and serial processing when capacity is unlimited at the individual element level. When we make that same comparison here we arrive at the same result: All else being equal, independent parallel processing is more efficient than serial processing.

Proposition 8.4: If the individual element processing times in serial and parallel models are identically distributed, then $G_{ST_2}(t) \geqslant F_{ST_2}(t)$ for all $t>0$.

Proof: Suppose the individual item processing time density function in the two models is $k(t)$. Then $G_{ST_2}(t)=K(t)$ and

$$F_{ST_2}(t)=\tfrac{1}{2}K(t)+\tfrac{1}{2}[K(t)*k(t)]$$

The proof of Proposition 8.3 shows that these two functions are ordered as the result requires. □

In concluding this section, there is one more comparison we might be interested in making, and that is between self-terminating and exhaustive search strategies. Intuition certainly suggests that a self-terminating search should be more efficient, and in this case our intuition turns out to be correct.

Proposition 8.5: If capacity is unlimited at the individual element level, $F_{ST_2}(t) \geqslant F_{EX_2}(t)$ for all $t>0$ and $G_{ST_2}(t) \geqslant G_{EX_2}(t)$ for all $t>0$.

Proof: We again assume the Proposition 8.1 definition of unlimited capacity at the individual element level.

(Serial proof) Under this definition the result is true if

$$\tfrac{1}{2}F(t)+\tfrac{1}{2}[F(t)*f(t)] \geqslant F(t)*f(t) \quad \text{for all } t>0$$

or equivalently if $F(t) \geqslant F(t)*f(t)$ for all $t>0$. This ordering was verified in Proposition 8.1.

(Parallel proof) The parallel result follows because $G(t) \geqslant G^2(t)$ for all $t>0$. □

All else being equal, self-termination is a more economical processing strategy than an exhaustive search, just as parallel processing is more economical than serial processing. Of course, the fact that a given processing strategy is analytically more economical is no guarantee it will (or can) be adopted by the processing organism.

It should be emphasized that, so far, all the results of this chapter assume unlimited capacity at the individual element level. Different capacity assumptions could lead to different conclusions. This point is perhaps obvious, but nevertheless is very important. For instance, in Proposition 8.3 we concluded that parallel models are unlimited capacity at the self-terminating level, given that they are unlimited capacity at the individual element level. If we had instead assumed individual element *limited* capacity – that is, if we had assumed that increasing the processing load decreases the individual element processing rates – then our conclusions about unlimited capacity at the self-terminating level would no longer follow. Instead we would conclude that parallel self-termination was limited capacity. Similarly, if individual element processing were supercapacity, we would find supercapacity at the self-terminating level in the presence of independent parallel processing. Analogous (but not equivalent) statements can, of course, be made about serial models.

Capacity at the individual element level

The term *unlimited capacity at the individual element level* has been a crucial one throughout this chapter. Whenever we used this phrase, we assumed that all individual element processing times were identically distributed for all processing loads and for all serial positions. Clearly this sounds like an example of unlimited capacity, but it also appears to be overly restrictive. Suppose the individual item processing times depend on serial position as well as processing load. For instance, in a parallel model suppose $G_{1,2}(t)> G_{1,1}(t)>G_{2,2}(t)$ for all $t>0$. Thus when the load is one element the processing rate is greater than the rate on position 2 when the load is two elements but less than the rate on position 1. Should we designate capacity limited, unlimited, or super in this situation?

As it turns out, we are faced with a dilemma when we try to generalize our capacity definitions: There does not appear to be any single definition that guarantees orderings for all interesting levels (e.g., exhaustive, self-terminating, and minimum completion times). Therefore, our strategy will be to define capacity in a way that relates directly to the distribution function at the level of interest and at the same time relates to capacity at the individual element level. This will have the advantage of providing distribution orderings at the appropriate level and yet assign a capacity category at the individual element level. A relatively minor price that will be exacted is that the definitions at two different levels will not be identical except when serial position effects are absent. However, we do win the generality of serial position asymmetries with this approach.

Throughout the rest of this chapter we shall concentrate (although not exclusively) on parallel models, thus reflecting our belief that the capacity issue is more naturally discussed in parallel processing systems (see Chapter 4). To ease the computational burden, all parallel models in this chapter assume independence of total completion times.

We can now present our definitions of capacity at the individual element

level. Each of these is in the form of an ordering on means of distribution functions. The means are taken over the different serial positions, and the orderings are in terms of the processing load on the system. The definitions differ in the type of mean employed. For instance, when search is self-terminating, arithmetic means are chosen.

Definition 8.1: When search is self-terminating, capacity is *limited* at the individual element level when

$$\frac{1}{n-1}\sum_{i=1}^{n-1} G_{i,n-1}(t) \geqslant \frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t) \quad \text{for all } t>0$$

and it is *super* when

$$\frac{1}{n-1}\sum_{i=1}^{n-1} G_{i,n-1}(t) \leqslant \frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t) \quad \text{for all } t>0$$

Each of these inequalities is assumed to be strict for at least one value of t and to hold for all values of n. If neither inequality is satisfied or if both are, capacity is said to be *unlimited.* □

Recall that the exhaustive distribution function involves a product of the separate distribution functions in independent parallel processing. It is therefore natural to employ the geometric mean as a way of expressing capacity at the level of the individual element when search is exhaustive. This approach leads to the following definition.

Definition 8.2: When search is exhaustive, capacity is *limited* at the individual element level when

$$\left[\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/(n-1)} \geqslant \left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{1/n} \quad \text{for all } t>0$$

and it is *super* when

$$\left[\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/(n-1)} \leqslant \left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{1/n} \quad \text{for all } t>0$$

Each of these inequalities is assumed to be strict for at least one value of t and to hold for all values of n. If neither inequality is satisfied or if both are, capacity is said to be *unlimited.* □

Finally, it will be instructive to generate a definition that readily applies to the first stage of processing or the minimum completion time, although the definition will see little service in the remainder of the chapter. The minimum completion time is observable, for example, when all elements to be searched match the target.

Definition 8.3: Capacity is *limited* in minimum time processing at the indi-

$$\left[\prod_{i=1}^{n-1} \bar{G}_{i,n-1}(t)\right]^{1/(n-1)} \leqslant \left[\prod_{i=1}^{n} \bar{G}_{i,n}(t)\right]^{1/n} \quad \text{for all } t>0$$

and it is *super* when

$$\left[\prod_{i=1}^{n-1} \bar{G}_{i,n-1}(t)\right]^{1/(n-1)} \geqslant \left[\prod_{i=1}^{n} \bar{G}_{i,n}(t)\right]^{1/n} \quad \text{for all } t>0$$

Each of these inequalities is assumed to be strict for at least one value of t and to hold for all values of n. If neither inequality is satisfied or if both are, capacity is said to be *unlimited.* □

Note that here again we see a product, but this time in the survivor functions rather than the distribution functions. Several facts follow immediately from these definitions. First, note that limited capacity defined, say, for self-terminating processing does not, in general, imply limited capacity for exhaustive or first-stage processing. In all these definitions capacity relates to the individual element but is simply expressed differently for the three types of processing (equivalently, for the three types of stopping rules). The three definitions do turn out to be equivalent in the absence of serial position effects; that is, when $G_{i,n}(t)=G_{j,n}(t)$ for all $i, j \leqslant n$.

It may be observed that the definitions of unlimited capacity include distributions that cross over as well as those that are identical. In the self-terminating models, an example of a crossover occurs when

$$\frac{1}{n-1}\sum_{i=1}^{n-1} G_{i,n-1}(t) < \frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t) \quad \text{for } t<t'$$

and

$$\frac{1}{n-1}\sum_{i=1}^{n-1} G_{i,n-1}(t) > \frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t) \quad \text{for } t \geqslant t'$$

In a sense, capacity is super for $t<t'$ and limited for $t \geqslant t'$. Such an instance might occur, for instance, if mean RT is unchanged but RT variance increases with processing load. Another possibility is that the self-terminating distribution functions are in fact identical but that sampling error causes one or more crossovers. In any event, we choose to include models that behave like this in the general category of "unlimited capacity."

The special case, in which no crossover occurs, satisfies

$$\frac{1}{n-1}\sum_{i=1}^{n-1} G_{i,n-1}(t) = \frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t) \quad \text{for all } t>0$$

We will refer to this important model as the *standard unlimited capacity model.* Clearly, in either type of unlimited capacity, it seems reasonable to say that overall, capacity is not increasing or decreasing at the (average) individual element level as the load changes, which is what we intuitively mean by unlimited capacity.

In the next section we consider a proposed test of self-terminating search

much as we have just been discussing. Because this test actually relates to several processing issues besides self-termination, and especially to capacity, it was not discussed in detail in the preceding chapter on self-terminating vs. exhaustive processing. Much of the above machinery will be brought into play in the following developments.

A proposed test of the self-terminating hypothesis

A recent development based upon distribution orderings was presented by Sternberg in 1973. He desired a test of self-termination that would be valid in a wide variety of contexts. In its most general formulation (Sternberg 1973, Appendix A1.1), a set of stimulus elements occupies a set of states of accessibility. Assumptions are postulated about the usually unobservable processing time distribution functions associated with these states and inferences drawn about observable RT distribution functions. Our investigation will proceed on the basis of target-present trials containing a single target vs. the target-absent trials with no target present. Vorberg (1981) has recently analyzed an extension of this design to tasks including $r>1$ targets on target-present trials; the models studied are serial (see also Chapter 7).

Sternberg refers to these states as being analogous to positions in a memory buffer, some of which are more easily reached than others. The most easily reached positions (i.e., most accessible) are associated with shorter processing times. Another possible interpretation of the states is as a set of processing locations, analogous to those we refer to throughout the book.

Sternberg makes four specific assumptions about the processing system and the set of internal states. First, he assumes that the states can be linearly ordered in terms of a weak ordering imposed by their distribution functions. In other words, with n distinct states he assumes that all states can be indexed so that whenever $i \leqslant j$ then $R_i(t) \geqslant R_j(t)$ for all $t>0$, where $R_i(t)$ is the probability (distribution function) that state i is accessed by time t. We say that state i is more accessible than j if the total completion time of an element in state i is less than the total completion time when that element is in state j (as measured by an ordering on the distribution functions). This first assumption, referred to as the *accessibility ordering* (AO) assumption, ensures that we can always completely order the set of states according to their accessibility.

A second assumption, called the *tight packing* (TP) assumption, states that the more accessible states are always filled first. For instance, with three elements to be processed, they must occupy the three most accessible states. Tight packing says that these are the states indexed 1, 2, and 3. Alternatively, if we know the nth state is filled, TP implies states $1, 2, \ldots, n-1$ are also filled.

Third is the *nested states* (NS) assumption, which says that accessibility, as defined on the distribution functions $R_i(t)$, is independent of the number of occupied states. Thus, the distribution function $R_i(t)$, which governs the total processing time of an element in state i, is the same whether there are i states filled or $i+j$ states filled.

This assumption of nested states (NS) is very similar to the assumption of pure insertion, first proposed by Donders (1868) more than a century ago. Pure insertion is, undoubtedly, among the oldest of the RT assumptions. It states that given an experimental task, another can be found, requiring all of the processing stages of the first plus one additional stage. This additional stage is said to have been inserted into the processing chain. An important requirement here is that the inserted stage have no effect on any of the other stages. They unfold in precisely the same fashion in both tasks. Thus if the nth stage is the inserted one, pure insertion requires that the duration of the $(n-1)$st stage does not depend on the existence of the nth stage. This is exactly the requirement imposed by NS.

The assumption of pure insertion has been an unpopular one for many years, partly because it is largely untestable with mean statistics, which is virtually the only level at which the assumption has been investigated. Recently however, Ashby and Townsend (1980) and Ashby (1982a) extended the assumption to the distributional level and found that in so doing, rigorous tests are indeed possible.

Within the present finite state model, however, the NS assumption seems to depend on the condition of tight packing (TP), in the sense that it is unclear what NS means if TP is not guaranteed. To see this, consider a serial system that always searches through the states in the order $1, 2, \ldots, n$. This strategy guarantees AO. With TP, $R_i(t)$ will always include the processing time of i elements, no matter how many states greater than i are filled; hence NS makes sense. However, without TP, $R_i(t)$ might be the distribution function for the processing time of between 1 and i elements. In this case the NS assumption seems questionable. A special case where this might make sense is if even empty states (cells, etc.) are checked as well as occupied ones.

The fourth assumption is that processing is self-terminating, and it is this assumption that Sternberg hoped to test. Clearly these assumptions define a class of models. Thus, empirical falsification of any theorems derived from these four assumptions must be related to the set as a whole, and not just to self-termination. Obviously, a test of the self-terminating hypothesis must assume that the first three assumptions are much weaker than the fourth (self-termination), so that falsification of the model would only very rarely occur through failure of one or more of AO, TP, and NS. If this were true, when violations did occur, they could be confidently attributed to a failure of self-termination. Since the strength of the three assumptions AO, TP, and NS is so crucial to the test, it will behoove us to examine them in more detail.

We begin by discussing experimental paradigms in which the observer searches for the presence or absence of a single critical or target element that may or may not be contained in an available stimulus array (e.g., memory scanning paradigms, visual search tasks). The self-terminating assumption is, of course, crucial only on those trials on which the target is in the display (target-present trials). When the target is not in the display (target-absent trials) processing is assumed to be exhaustive.

When the target is in buffer position i, the processing time distribution

function is $R_i(t)$. Therefore, if the experimental placement of the target in the display is random with regard to whatever might correspond to the Sternberg states and with regard to the variables affecting the observer's positioning of items in the states (e.g., temporal position in auditory presentation, spatial position in visual presentation), then it is easily shown that the self-terminating distribution function when n elements are in the display, $Q_n(t)$, can be written as

$$Q_n(t)=\frac{1}{n}\sum_{i=1}^{n} R_i(t)$$

Sternberg (1973) has called this the invariance property to emphasize that $Q_n(t)$ "depends only on the set of occupied states and not on the occupation probabilities."

With the foregoing model, we can now prove the two important properties that Sternberg proposed as tests of self-termination. The Q terms in the following expressions are the empirical (observed) RT distribution functions for target-present trials.

Proposition 8.6 (long-RT property): If AO, TP, and NS hold and if search is self-terminating, then $Q_{n-1}(t)\geqslant Q_n(t)$ for all $t>0$.

Proof: From the assumption of AO we know that $R_n(t)\leqslant R_i(t)$ for all $i\leqslant n$ and for all $t>0$. Therefore, for all $t>0$

$$R_n(t)\leqslant \frac{1}{n-1}\sum_{i=1}^{n-1} R_i(t)=Q_{n-1}(t) \tag{8.1}$$

Now

$$Q_n(t)=\frac{1}{n}\sum_{i=1}^{n} R_i(t)=\frac{1}{n}\left[\sum_{i=1}^{n-1} R_i(t)+R_n(t)\right]=\frac{1}{n}\left[(n-1)Q_{n-1}(t)+R_n(t)\right]$$

But from Eq. 8.1 we know that $R_n(t)\leqslant Q_{n-1}(t)$, and so

$$Q_n(t)\leqslant \frac{1}{n}\left[(n-1)Q_{n-1}(t)+Q_{n-1}(t)\right] \quad \text{for all } t>0$$

and finally that $Q_n(t)\leqslant Q_{n-1}(t)$ for all $t>0$. □

Therefore, an ordering on the unobservable $R_i(t)$ establishes an ordering on the observable RT distribution functions. The long-RT property thus implies that the probability that reaction time is less than any particular time t when any number j of states are filled is less than that probability (for the same t) when i states are filled, where $j>i$. The intuition Sternberg had in mind when naming this property is that by adding an element to the display, say the $(i+1)$st, we are, because of TP, filling the $(i+1)$st state, which according to AO is less accessible than states $1, 2, \ldots, i$. Therefore, trials with $i+1$ items in the display should generate at least as many long RTs as trials with only i items.

It is important to notice however, that AO, TP, and NS are sufficient to prove the long-RT property; self-termination is not required. This is trivially proven.

Proposition 8.7: If AO, TP, and NS hold and if search is exhaustive, then $Q_{n-1}(t)\geqslant Q_n(t)$ for all $t>0$.

Proof: When processing is exhaustive, $Q_n(t)=R_n(t)\leqslant R_{n-1}(t)=Q_{n-1}(t)$, by the axioms of AO, TP, and NS. □

Thus the long-RT property cannot by itself be used to test between self-terminating and exhaustive search strategies. Both classes of models predict it to be true. Instead, it can be viewed as a test of the set of axioms: accessibility ordering (AO), tight packing (TP), and nested states (NS). If the long-RT property is found not to hold, then one or more of these assumptions must be false.

We now state and prove the second RT property that Sternberg proposed as a test of self-termination.

Proposition 8.8 (short-RT property): If TP and NS hold and if search is self-terminating, then

$$nQ_n(t)\geqslant (n-1)Q_{n-1}(t) \quad \text{for all } t>0$$

Proof: It is obvious that

$$\sum_{i=1}^{n} R_i(t)\geqslant \sum_{i=1}^{n-1} R_i(t) \quad \text{for all } t>0$$

and therefore

$$n\left[\frac{1}{n}\sum_{i=1}^{n} R_i(t)\right]\geqslant (n-1)\left[\frac{1}{n-1}\sum_{i=1}^{n-1} R_i(t)\right] \quad \text{for all } t>0$$

Thus

$$nQ_n(t)\geqslant (n-1)Q_{n-1}(t) \quad \square$$

This property states that for different values of j, $j\times Q_j(t)$ cannot be greater than $n\times Q_n(t)$ when $n>j$. Obviously, this inequality becomes trivially true when t is large enough that $Q_n(t)$ is close to 1; therefore, the only interest in the property is for small t. Intuitively, no matter how many elements are contained in the display, there should always occur trials on which the target will be the first element completed. If search is self-terminating, processing can end here and therefore all display sizes should generate some equally short reaction times.

Although the AO assumption was necessary to prove the long-RT property, it was not used in the proof of the short-RT property. Thus, self-termination

and NS suffice for the long-RT property. The two properties, taken in conjunction, induce a set of bounds on the $Q_i(t)$ such that for any $j>i$,

$$Q_j(t) \leqslant Q_i(t) \leqslant \frac{j}{i} Q_j(t)$$

Notice that both the short- and long-RT properties make predictions about the ordering of distributions when the processing *load* is changed. As the reader no doubt anticipates from our earlier comments, we believe that such predictions are intimately related to the capacity structure of the system regardless of their value in determining whether processing is self-terminating or exhaustive, or whether it is parallel or serial.

Relationship of Sternberg's four assumptions to short- and long-RT properties

Earlier we stated that Sternberg proposed the short- and long-RT properties as tests of self-termination. Thus, failure of either property was taken as evidence against a self-terminating search. However, we also noted that the empirical failure of any property of the model (e.g., the short- and long-RT properties) must be related to *all* the assumptions. We might then investigate more formally exactly what a violation of one or both of these properties really does imply about the entire set of assumptions and in particular the assumption of self-termination. Before we begin, we illustrate in Fig. 8.1 what the success and failure of these two properties might look like in the case when there are one or two elements in the display.

Long-RT property fails and short-RT property holds

Violation of the long-RT property implies that $Q_{n-1}(t') < Q_n(t')$ for *some* $t'>0$, whereas support of the short-RT property occurs when

$$Q_{n-1}(t) \leqslant \frac{n}{n-1} Q_n(t) \quad \text{for } \textit{all } t>0$$

Note that if the long-RT property fails everywhere, so that $Q_{n-1}(t) < Q_n(t)$ for all t (i.e., the ordering is reversed from what the long-RT property predicts), then the short-RT property must be satisfied, since if $Q_{n-1}(t)$ is less than $Q_n(t)$, it is necessarily less than $[n/(n-1)]Q_n(t)$. Hence, all we can conclude with certainty is that either AO, TP, or NS or some combination of these is at fault. It is important to keep in mind here and in what follows that the success of a property (in the present case, the short-RT property) does not imply the truth of its sufficient conditions (here, TP, NS, and self-termination). Perhaps the most attractive of these possibilities would be that AO fails while TP, NS, and self-termination hold, since AO was not one of the assumptions required to prove the short-RT property. However, there is no logical necessity for this conclusion. Note again that the exhaustive vs. self-terminating issue is not touched on when an experiment produces $Q_{n-1}(t) < Q_n(t)$.

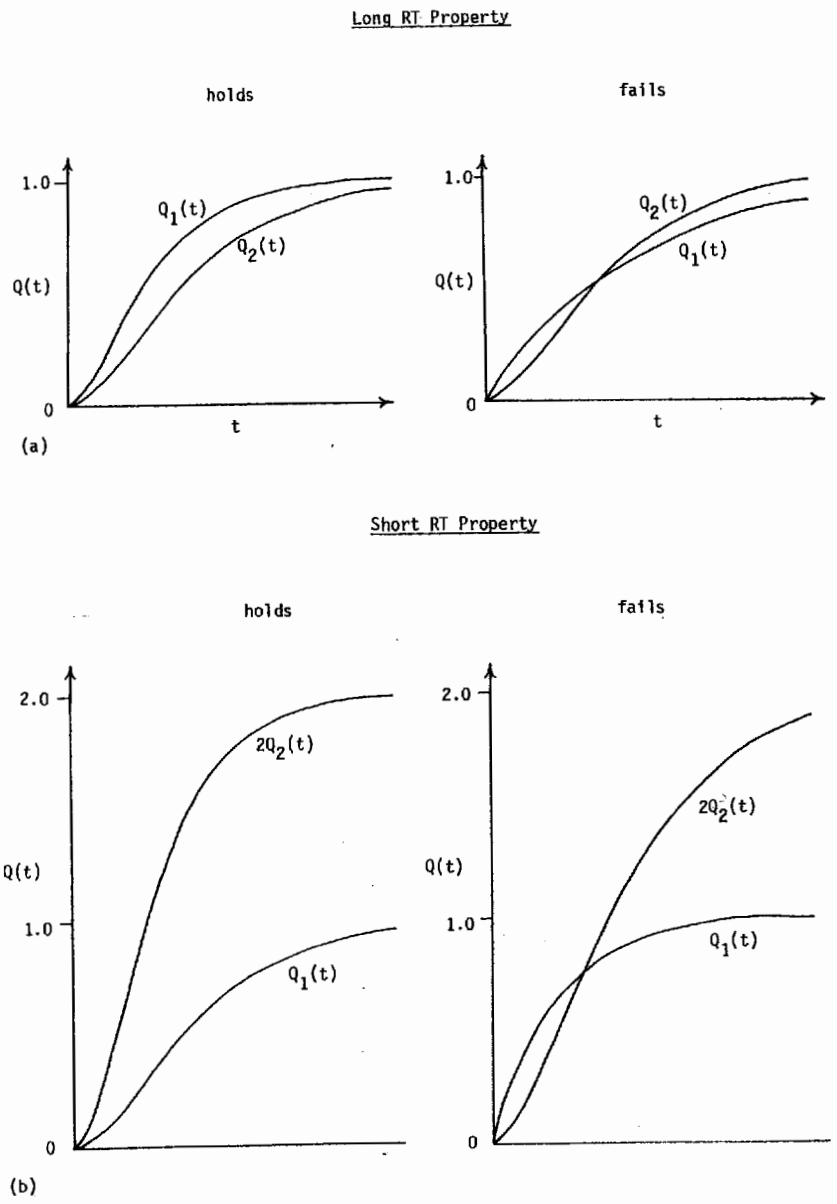


Fig. 8.1. What the success and failure of the long-RT property (a) and the short-RT property (b) look like when the processing load is one and two elements.

As an example of how the short-RT property can hold but not the long, suppose that TP holds and consider the case of Fig. 8.1 where display sizes are 1 and 2. A violation of the long-RT property for some time t implies, for that time, that $Q_1(t) < Q_2(t)$ and therefore that

$$\tfrac{1}{2}[R_1(t) + R_2(t)] > R_1(t)$$

Here $R_1(t)$ is the same on both sides of the inequality because of NS. If we drop this assumption, the above equation may be rewritten as

$$\tfrac{1}{2}[R_{1'}(t) + R_2(t)] > R_1(t)$$

where $R_{1'}(t)$ may or may not equal $R_1(t)$. Simplifying, we see that a failure of the long-RT property at time t implies $R_{1'}(t) + R_2(t) > 2R_1(t)$, so that either $R_{1'}(t) > R_1(t)$ or $R_2(t) > R_1(t)$ (or both). In the first case NS is violated, and in the second AO is violated. As pointed out, knowledge that the short-RT property holds does not help us decide which of NS or AO is faulty. The short-RT property states that $2Q_2(t) \geqslant Q_1(t)$ and therefore that $R_{1'}(t) + R_2(t) \geqslant R_1(t)$. This condition is implied by a failure of the long-RT property. In one sense this is not surprising since, it will be recalled, the short-RT property states that both $Q_2(t)$ and $Q_1(t)$ should generate some equally short RTs. We have just seen that a possible failure of the long-RT property is if $R_{1'}(t) > R_1(t)$, that is, if the minimum RT is stochastically *less* with two elements in the display. Certainly if the minimum RT decreases, we should expect the short-RT property to hold.

Short-RT property fails and long-RT property holds

This occurrence gives the ordering

$$Q_n(t) \leqslant \frac{n-1}{n} Q_{n-1}(t)$$

for one or more values of time t. At all other times the predicted order occurs. Notice that when the short-RT property fails over some interval, the long-RT property is implied. Clearly TP, NS, or the self-termination assumption must be wrong (or more than one of these). Of course, there are a very large number of non-self-terminating (e.g., exhaustive) models possessing the characteristics of TP, AO, and NS that will yield this type of result. But unless one is willing to place some kind of measure on the various assumptions that correspond to their likelihood in reality, it is impossible to rule out other alternatives (such as a system with self-termination, TP, and AO, but not NS), with only the present empirical information to go by.

Let us make some of these thoughts a little clearer by way of our $n=2$ example. Again, for simplicity we assume TP holds. In this case we know that $Q_1(t) \geqslant Q_2(t)$ for all $t > 0$ and $2Q_2(t') < Q_1(t')$ for *some* $t' > 0$, and therefore if processing is self-terminating,

$$R_1(t) \geqslant \frac{R_{1'}(t) + R_2(t)}{2} \quad \text{and} \quad R_{1'}(t') + R_2(t') < R_1(t')$$

Notice that we again specify $R_1(t)$ differently depending upon how many states are filled, in case NS is false. From the failure of the short-RT property we see that $R_2(t') < R_1(t') - R_{1'}(t')$ for some $t' > 0$. Now $R_2(t)$ must always be greater than or equal to zero, and therefore $R_1(t') > R_{1'}(t')$ for some $t' > 0$; NS fails because of an apparent capacity limitation. Note also that AO may or may not hold. At present it is unclear how strong the assumption of NS really is, for the severity of the assumption depends upon one's definition of the states of the system. If we assume the state refers to the number of elements processed by the system, then NS seems reasonable. The system may still slow down or speed up with processing load changes, but the processing time of the first j elements may not depend on these changes. If, on the other hand, we assume the state of the system refers to the various serial positions of the stimulus elements in the display, then the validity of NS is more tenuous.

Thus, in serial systems in which elements are always placed in the same positions and only one order of processing is ever employed, it is eminently reasonable to expect the NS axiom to be satisfied. Even if more than one processing order and one set of positions are used, NS may be correct if the processing distributions depend at most on the number of elements processed. Consider, for instance, multisymbol visual displays. If the displays always begin at the same retinal location and extend to the right as elements are added, and if processing rates on the succeeding positions do not change as the other elements are added, then NS seems acceptable. However, if the displays are centered with regard to the fovea, or if lateral interference occurs among the various elements, it seems very unlikely that a NS assumption would hold under this interpretation of state.

What happens if processing is exhaustive? Clearly, processing might be exhaustive and a number of other assumptions may be false. However, for instructive purposes, suppose the others are true and again let $n=1$. The long-RT property results, of course, but the short-RT property may or may not be found. Exhaustive processing and the failure of the short-RT property yields the relation $2Q_2(t) = 2R_2(t) < R_1(t) = Q_1(t)$ for some value of t. Thus, we observe that the AO axiom holds in spades: There is at least one processing time for which the probability that the element in the first position is completed is more than twice as great as the same probability for the element in the second position. In a serial system the latter is equivalent to the probability that both elements are finished by that time. In a parallel system, of course, on any single trial the first-position element may or may not be completed when the second-position element is done. In either system, the conclusion is that the system is quite limited in capacity.

On the other hand, we should be aware that exhaustive processing does not by itself automatically imply that the short-RT property will be false. When it

does hold and processing is exhaustive, then, for $n=2$, $R_1(t) \leqslant 2R_2(t)$, which does not violate AO, NS, or TP.

Both properties fail

A failure of both properties implies that

$$Q_{n+1}(t) > Q_n(t) > \frac{n+1}{n} Q_{n+1}(t) \quad \text{for some } t>0$$

Note that this simultaneous failure is impossible at the same time t. That is, if $Q_n(t') < Q_{n+1}(t')$ for some $t=t'$, then obviously at that same $t=t'$ it must be that

$$Q_n(t') < \frac{n+1}{n} Q_{n+1}(t')$$

Thus if both properties fail, they must do so at different times.

Both properties hold

In the event that neither of the predictions are falsified, we have garnered some support for the axioms TP, NS, AO, and self-termination. We must keep in mind, though, that there may be other equally intuitive models that are able to make these two predictions. If processing is exhaustive, then the long-RT property is implied by AO, TP, and NS, but the short-RT property is not implied, although it may be true. If search is exhaustive and the two properties are found to hold, then since $Q_n(t)=R_n(t)$ in exhaustive models, the following ordering is imposed on the $R_i(t)$:

$$R_n(t) \leqslant R_{n-1}(t) \leqslant \frac{n}{n-1} R_n(t) \quad \text{for all } t>0$$

The left-hand part of the inequality is telling us that the positions farther down in the buffer take longer to get to (i.e., are stochastically less accessible), which we know already from AO in conjunction with the other assumptions. However, the right-hand side of the inequality says that the ratio of R_{n-1} to R_n can never be greater than the ratio of n to $n-1$ for any value of processing time t. This puts a clear constraint on how limited capacity can be by limiting the decrease in the distribution function to

$$R_n(t) \geqslant \frac{n-1}{n} R_{n-1}(t)$$

There are, of course, other systems that can predict the short- and long-RT properties but that violate some of the above assumptions. For example, notice that even if processing is self-terminating and both predicted RT properties hold, the other three assumptions AO, TP, and NS are not implied.

Thus, in our $n=2$ example, if we assume TP and we allow for a change in R_1 when going from $n=1$ to $n=2$ our two properties give us

$$\frac{R_{1'}(t)+R_2(t)}{2} \leqslant R_1(t) \leqslant R_{1'}(t)+R_2(t)$$

It is reasonably apparent that we can permit $R_1(t) > R_{1'}(t)$ (violating NS) and $R_2(t) > R_1(t)$ (violating AO) for at least some t without violating the predicted ordering.

In closing this section, we might summarize our findings as follows. First, in any given experiment, it may be difficult to detect just which axioms are true, if all we have to base our judgments on are tests of the long- and short-RT properties. With regard to the joint assumptions of AO, TP, and NS, a failure of the long-RT property will be particularly damaging since both self-terminating as well as exhaustive models predict it. This being the case, the short-RT property is the one more relevant to testing self-terminating vs. (some) exhaustive models. However, without additional information, there may be no reason to suspect self-termination rather than another assumption when a short-RT property violation is found. We now turn to a discussion of how general serial and parallel models might interpret the three major assumptions (AO, TP, and NS) defining this class of finite state models.

Serial and parallel interpretations of Sternberg's finite model

Serial interpretations

Before proceeding further, let us examine the possible serial and parallel interpretations of the above assumptions. The serial interpretations are straightforward. One possibility is to identify the states with the serial positions in the display, but perhaps the most general approach is to simply view the states as registers in some unnamed buffer (e.g., a short-term memory buffer). These registers might be filled according to the serial positions of the to-be-processed elements. Thus a states-as-serial-positions interpretation is a special case of this one. If we assume each register can contain at most one element and that register or state i is always searched before state j if $j>i$, then the state will refer to the number of elements processed by the system. Thus the index simply refers to the order of processing, whatever it is, 1 being the index of the position searched first, 2 of the one examined second, and so on.

In this interpretation, AO implies that the distribution function for processing one element is always greater than or equal to the distribution function for two elements. In other words, the probability that a system has completed processing one element by any time t can never be less than the corresponding probability of completing two elements, an apparently weak assumption. If the processing times are independent on the separate posi-

tions, then the density function $r_i(t)$ [corresponding to the distribution function $R_i(t)$] is a convolution of the i independent serial densities $f_j(t)$, $j=1,2,\ldots,i$ and the AO assumption immediately follows from TP and NS, both of which are eminently plausible in the context of this serial interpretation.

To derive AO, we need only follow the serial argument in Proposition 8.1. Thus note that by integration we can write

$$\begin{aligned} R_i(t) &= R_{i-1}(t) * f_i(t) \\ &= \int_0^t R_{i-1}(t-t')f(t')\,dt' \\ &\leqslant R_{i-1}(t)\int_0^t f(t')\,dt' \\ &\leqslant R_{i-1}(t) \quad \text{for all } t>0 \end{aligned}$$

which proves AO.

This serial interpretation naturally explains such phenomena as serial position effects. If an observer always processes a display of elements in the same order, then on different trials the state index will refer to given serial positions within the display, as well as to processing order. The processing time of the system for each serial position of the critical or target element within the display will always be associated with one and the same distribution function. The degree to which these distribution functions differ will determine the size of the serial position effects and the processing order that the system selects will determine the shape of the serial position curves.

Thus, the assumptions of AO, TP, and NS give a serial model that conforms closely to the usual standard serial interpretations. For example, the assumptions lead to limited capacity at the exhaustive and the self-terminating levels but unlimited capacity at the level of the first intercompletion time (i.e., minimum RT). This is exactly the prediction of standard serial models with unlimited capacity at the individual element level, but of course includes other models as well.

Parallel interpretations

The independent parallel interpretation is slightly more difficult to come by. The most natural interpretation might be to identify the states with the serial positions of the items in the display. In this case, we can let the second subscript (n) in the parallel distribution function $G_{i,n}(t)$, stand for the total number of items to be processed and the first subscript stand for the particular serial position of an item $(i=1,2,\ldots,n)$. Under this interpretation, $G_{i,n}(t)$ is the probability that the item in position i or state i is completed by time t, given there are n items in the display.

Recall that AO requires the processing time of the item in state (i.e., position) i to be stochastically less than (or equal to) the processing time of the item in state $i+1$. Thus, AO might be interpreted as an ordering on capacity, so that successively less capacity is allotted to serial positions indexed by a greater number. For a fixed load this restriction is trivial since we can always order a set of items by the amount of capacity they are allocated [remember AO implies the *weak* order $R_i(t) \geqslant R_{i+1}(t)$].

With a visual search through a horizontal linear array, AO might correspond to the diminishing of attention or processing capability with distance from the fovea. Thus the more central stimulus elements would tend to be assigned to the higher states (assuming state 1 is the highest). Now suppose the processing load – or equivalently, the number of display elements – is increased. NS now becomes crucial since it says that although the overall capacity (as reflected in mean RT, for example) may decrease as n increases, the capacity for any particular state or position cannot change with n. In other words, $G_{i,n}(t)=G_i(t)=R_i(t)$.

This interpretation is, however, limited in another respect. Recall that the invariance property requires $Q_n(t)$ to be an equally weighted sum of the $R_i(t)$; that is,

$$Q_n(t)=\frac{1}{n}\sum_{i=1}^{n} R_i(t)=\frac{1}{n}\sum_{i=1}^{n} G_i(t)$$

Suppose now that a critical or target element is always processed much faster than a nontarget so that with parallel processing it is almost always completed first no matter what serial position it occupies within the display. Thus, within the terminology of our parallel interpretation we would say that target elements are consistently allocated a disproportionately larger relative share of capacity and therefore are more often "assigned" to states indexed by a lower number than are nontargets. Obviously, with a self-terminating search the invariance property will not now hold. Distribution functions indexed by lower numbers would then possess a greater weight in the determination of $Q_n(t)$. Analogous arguments can be put forth if targets are processed slower than nontargets. In fact, the invariance property will not hold if processing rate depends on element identity in any fashion whatsoever. The reader may note that this problem is related to the issue of parallel–serial discriminability when processing rate differences exist for target and nontarget elements (see Chapters 6, 7, and 13).

This discussion might lead us to expect that not all of the independent parallel models introduced at the beginning of this chapter will satisfy the first three of Sternberg's assumptions (i.e., AO, TP, and NS) under the present interpretation of the states. This is indeed the case (Townsend 1974b). It is easy to show that very limited capacity, as well as all unlimited and supercapacity models, are incompatible with at least one of Sternberg's assumptions.

First recall that NS demands that $G_{i,n}(t)$ must be independent of the processing load n, so that $G_{i,n}(t)=G_i(t)$ for all values of n. Meanwhile, AO guarantees that the individual element distribution functions are ordered according to position, that is, that

$$G_1(t)\geqslant G_2(t)\geqslant\cdots\geqslant G_n(t) \quad \text{for all } t>0$$

Tight packing (TP), on the other hand, asserts that n items will always fill the first n positions so the G_i notation is appropriate for each trial and varying n.

Thus when AO, NS and TP hold we know

$$\frac{1}{j}\sum_{i=1}^{j} G_i(t)\geqslant\frac{1}{n}\sum_{i=1}^{n} G_i(t) \quad \text{for all } t>0$$

and for all $j<n$. Along with self-termination these imply

$$Q_j(t)\geqslant Q_n(t) \qquad \text{for all } t>0$$

This is the condition we earlier suggested as a definition of limited capacity at the individual element level. Thus, taken together, AO, NS, and TP imply limited capacity processing at the level of the individual element. As such, all unlimited capacity and all supercapacity independent parallel models must therefore be incompatible with one or more of the three assumptions.

This does not mean to say, however, that all limited capacity models satisfy the set of three assumptions – they do not. In other words, limited capacity is a necessary but not a sufficient condition for an independent parallel model to satisfy AO, NS, and TP. As an almost trivial example, suppose capacity is very limited so that, for instance, $G_{i,n}(t)=G^{jn}(t)$ for some distribution function $G(t)$. Then AO holds and TP can hold, but NS does not because $G_{i,n}(t)$ depends on n.

In addition to illustrating that not all limited capacity parallel models satisfy AO, NS, and TP, this particular model turns out to predict a violation of the short-RT property and so underscores the point we made earlier that not all self-terminating models predict the short-RT property to be true. Remember that when $n=2$, the short-RT property states that $2Q_2(t)\geqslant Q_1(t)$ for all $t>0$. With parallel self-terminating models this property becomes $G_{1,2}(t)+G_{2,2}(t)\geqslant G_{1,1}(t)$ for all $t>0$, and under the capacity limitations of the example we are considering, the inequality can be written as $G^2(t)+G^4(t)\geqslant G(t)$ for all $t>0$, which reduces to $G(t)+G^3(t)\geqslant 1$ for all $t>0$. There will always exist many t for which this statement is false. In fact, the inequality is violated for all $t\leqslant t'$, where t' has the property that $G(t')=.682$ [i.e., $t'=G^{-1}(.682)$]. Thus the short-RT property is severely violated in this example.

Of course, as we have seen, NS could be violated without necessarily violating the short-RT property – because of capacity limitations, for example. Obviously, AO and TP might also be violated in independent models, in which case the short- and long-RT properties might or might not be satisfied.

We may summarize the results of this section as follows: Under the interpretation of states as serial positions in an independent parallel system, AO, TP, and NS imply that capacity is moderately limited.

In the next section we leave the issue of the unobservable assumptions AO, NS, and TP and concentrate on their observable consequences, namely the long- and short-RT properties. Specifically we will ask two questions: (1) Which independent parallel models can (and cannot) predict these two properties? and (2) What can we infer about processing from some specific outcome of some empirical test of the two properties?

Since we will no longer be dealing specifically with the AO, NS, and TP assumptions, we need not concern ourselves with a specific parallel interpretation of the states. The two questions we will ask can be answered without an appeal to such an interpretation.

The short-RT property and independent parallel models

An investigation of which independent parallel models can predict the short- and long-RT properties is most illuminating when it incorporates a definition of capacity formulated at the level of the individual item processing times. We could construct capacity definitions on either the exhaustive or self-terminating completion times, depending on which stopping rule we are considering, but this strategy turns out to lead to generally less interesting results. Thus the following discussion will utilize our earlier definitions based on the arithmetic (with self-terminating search) and geometric (with exhaustive search) means of the individual item processing time distribution functions. We are neglecting the residual RT component from extraneous processes. So strictly speaking, one should subtract (deconvolve) an estimate of the residual time distribution before applying the following results empirically. However, the same basic conclusions appear to follow if the theorems are carried out inclusive of the residual component.

Using these capacity definitions, self-terminating models are especially easy to investigate because, in parallel models, the observable RT distribution function is precisely the arithmetic mean of the individual item processing time distribution functions,

$$Q_n(t)=\frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t)$$

With this in mind, we state the first result.

Proposition 8.9: All supercapacity and all standard unlimited capacity, independent parallel self-terminating models predict the short-RT property to be true.

Proof: It may be recalled that supercapacity, independent parallel self-terminating models are those for which

$$\frac{1}{n-1}\sum_{i=1}^{n-1} G_{i,n-1}(t) \leqslant \frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t) \quad \text{for all } t>0$$

If we relax the requirement that the inequality be somewhere strict, then the standard unlimited capacity model also satisfies this condition. But as noted above, these are just the observable RT distribution functions. Thus for independent parallel self-terminating models, standard unlimited capacity and supercapacity models are those for which $Q_{n-1}(t) \leqslant Q_n(t)$ for all $t>0$. Now this inequality implies

$$(n-1)Q_{n-1}(t) \leqslant nQ_{n-1}(t) \leqslant nQ_n(t) \quad \text{for all } t>0$$

And hence

$$(n-1)Q_{n-1}(t) \leqslant nQ_n(t) \quad \text{for all } t>0$$

which is the short-RT property. □

This is a very natural result. The short-RT property is meant to reflect the fact that, no matter what the load, all self-terminating models will generate some equally short RTs. With unlimited and supercapacity models, the capacity allotted to each individual item either increases or at least does not decrease as the load increases. Because the processing time of a parallel self-terminating model is determined entirely by the processing time of the target element, it makes sense that with unlimited and supercapacity models, as short or even shorter RTs will be generated when the processing load is increased and thus that the short RT property should hold.

In addition, we already know that some limited capacity self-terminating models predict the short-RT property since we showed that parallel models that are moderately limited in capacity are compatible with Sternberg's assumptions of AO, TP, and NS under the states-as-positions interpretation. As such they must predict the short-RT property to be true.

The general class of exhaustive models can also be investigated in this manner. As we have seen, because of independence, the exhaustive distribution function may be written as

$$Q_n(t) = G_{1,n}(t)G_{2,n}(t)\cdots G_{n,n}(t) = \prod_{i=1}^{n} G_{i,n}(t)$$

We may now state our second result.

Proposition 8.10: No limited capacity or standard unlimited capacity independent parallel exhaustive model can predict the short-RT property to be true.

Proof: When processing is parallel and exhaustive, the short-RT property

$$(n-1)Q_{n-1}(t) \leqslant nQ_n(t) \quad \text{for all } t>0$$

is equivalent to

$$(n-1)\prod_{i=1}^{n-1} G_{i,n-1}(t) \leqslant n\prod_{i=1}^{n} G_{i,n}(t) \quad \text{for all } t>0$$

Let us first consider the class of standard unlimited capacity models, namely, those for which

$$\left[\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/(n-1)} = \left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{1/n} \quad \text{for all } t>0$$

We may express the short-RT property as

$$n\prod_{i=1}^{n} G_{i,n}(t) = n\left[\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{n/(n-1)} \geqslant (n-1)\prod_{i=1}^{n-1} G_{i,n-1}(t)$$

which, after some simplification, may be rewritten as

$$\left[\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/(n-1)} \geqslant \frac{n-1}{n}$$

which is obviously false as t approaches 0.

The result now proceeds a fortiori in the case of limited capacity models. Or the reader may construct a proof analogous to the one for standard unlimited capacity models. □

This result seems to indicate that exhaustive models, like their self-terminating counterparts, have an easier time predicting the short-RT property to be true when they are supercapacity. Indeed, some supercapacity models can predict it to be true. However, it turns out that a very large class of exhaustive models can be shown incapable of predicting the property, no matter how super in capacity they are.

Proposition 8.11: No independent parallel exhaustive model without serial position effects can predict the short-RT property to be true if its individual element processing densities are continuous and differentiable about (in a half-open interval) the origin and if $g_n(0)$ is a nonzero finite value.

Proof: When processing is independent, parallel, and exhaustive and there are no serial position effects, the short-RT property becomes

$$(n-1)[G_{n-1}(t)]^{n-1} \leqslant n[G_n(t)]^n \quad \text{for all } t>0$$

Note that this relation is always an equality at the origin. Thus, if the derivative of the left side is greater than the derivative of the right side over some region $(0, \Delta t)$, then the short-RT property will be violated.

Evaluating the derivative of both sides of the above expression gives

$$(n-1)^2 g_{n-1}(t)[G_{n-1}(t)]^{n-2} \quad \text{vs.} \quad n^2 g_n(t)[G_n(t)]^{n-1}$$

We are particularly interested in comparing these two derivatives at the origin. Specifically, for a violation of the short-RT property we must show that

$$(n-1)^2 g_{n-1}(t)[G_{n-1}(t)]^{n-2} > n^2 g_n(t)[G_n(t)]^{n-1}, \quad 0 \leqslant t < \Delta t$$

where Δt is arbitrary. Now this inequality holds whenever

$$\left(\frac{n-1}{n}\right)^2 \frac{g_{n-1}(t)}{g_n(t)} > \frac{[G_n(t)]^{n-1}}{[G_{n-1}(t)]^{n-2}}$$

By the assumptions of the result, the left-hand side is greater than zero over some interval $(0, \Delta t)$, and therefore if we can show that

$$\lim_{t \to 0} \frac{[G_n(t)]^{n-1}}{[G_{n-1}(t)]^{n-2}} = 0$$

the proof will be complete. Applying L'Hospital's rule to this ratio $n-2$ times gives us

$$\lim_{t \to 0} \frac{[G_n(t)]^{n-1}}{[G_{n-1}(t)]^{n-2}} = \lim_{t \to 0} \frac{(n-1) g_n(t)[G_n(t)]^{n-2}}{(n-2) g_{n-1}(t)[G_{n-1}(t)]^{n-3}}$$

$$= \lim_{t \to 0} \frac{(n-1)![g_n(t)]^{n-2} G_n(t) + \sum_{i=1}^{2n-5} K_i}{(n-2)![g_{n-1}(t)]^{n-2} + \sum_{i=1}^{2n-5} H_i} = 0$$

under the conditions of the result because H_i and K_i represent terms containing the distribution functions $G_{n-1}(t)$ and $G_n(t)$, respectively, and so

$$\lim_{t \to 0} H_i = \lim_{t \to 0} K_i = 0 \quad \text{for } 1 \leqslant i \leqslant 2n-5 \quad \square$$

Note that the restriction $g_n(0) > 0$ guarantees that there is some nonzero probability density at $t = 0$. This implies that the probability of completing the ith element in the first instant after time $t = 0$ will be positive. Apparently this condition is enough to guarantee a violation in the short-RT property.

Many distributions satisfy this restriction on the $g(t)$ terms. Included in this class are all models with ordinary exponential processing [i.e., if $g_n(t) = v_n \exp(-v_n t)$, then $g_n(0) = v_n > 0$]. But the really striking thing is that this result holds for all capacity levels. No matter how fast v_n increases with n, these exhaustive models can still not predict the short-RT property, as the reader may wish to verify. In fact, letting v_n be a function of n *and* time $[v_n(t)]$ is also to no avail, thus underscoring the fact that Proposition 8.11 is not specific to exponential distributions.

Let us illustrate this result for the simple case in which the $g_{i,n}$ are uniform distributions. Suppose $g_1(t) = 1$ for $0 \leqslant t < 1$ and 0 elsewhere and $g_2(t) = 1/a$ for $0 \leqslant t < a$ and 0 elsewhere. In this case supercapacity results whenever $a < 1$. Now in this model, $G_1(t) = t$ for $0 \leqslant t < 1$ and $G_2(t) = t/a$ for $0 \leqslant t < a$. Thus the short-RT property can be written as

$$G_1(t) = t \leqslant 2\left(\frac{t}{a}\right)^2 = 2[G_2(t)]^2 \quad \text{for all } t > 0$$

which is true only when $t \geqslant a^2/2$. In other words, no matter how small a is, or equivalently, no matter how great the capacity increase is as the load is incremented by one item, the short-RT property will always be violated (so long as $a > 0$) in the interval $(0, a^2/2)$.

As it turns out, the continuity requirement of Proposition 8.11 is very important, and some interesting counterexamples can be constructed in cases when it is violated. For instance, suppose as before that g_1 is uniform over the interval $(0, 1)$, but now let $g_2(t) = at^{a-1}$ for $0 \leqslant t < 1$. Then $G_2(t) = t^a$ for t in this same interval. The short-RT property can now be written as $G_1(t) = t \leqslant 2t^{2a} = 2[G_2(t)]^2$ for all $t > 0$, or equivalently, $t^{1-2a} \leqslant 2$ for all $t > 0$. When $1 - 2a \geqslant 0$, then $t^{1-2a} \leqslant 1$ over the interval of interest (i.e., $0 \leqslant t < 1$), which implies the short-RT property will hold whenever $a \leqslant \frac{1}{2}$. The continuity condition and $g_n(0)$ positive and finite of Proposition 8.11 are violated in this example, however, since $\lim_{t \to 0} g_2(t) = \infty$. Thus, the example does not satisfy the assumptions of the proposition. It does, however, affirm that at least some parallel exhaustive models predict the short-RT property to be true, and so the property cannot logically be used to discriminate all parallel exhaustive from parallel self-terminating models.

The long-RT property and independent parallel models

We now move to an investigation of the long-RT property.

Proposition 8.12: If we consider the class of independent, parallel self-terminating models, then the standard unlimited capacity model and *all* models with *limited* capacity predict the long-RT property to be true, whereas *all super*capacity models predict it to be false.

Proof: Limited capacity, independent parallel self-terminating models are defined by

$$\frac{1}{n-1} \sum_{i=1}^{n-1} G_{i,n-1}(t) \geqslant \frac{1}{n} \sum_{i=1}^{n} G_{i,n}(t) \quad \text{for all } t > 0$$

In terms of the observable RT distribution function, this inequality becomes $Q_{n-1}(t) \geqslant Q_n(t)$ for all $t > 0$, which is precisely the long-RT property.

Supercapacity models are defined by the condition $Q_{n-1}(t) \leqslant Q_n(t)$ for all $t > 0$, with a strict inequality holding for some time t. This last requirement guarantees a long-RT property violation.

The standard unlimited capacity models trivially satisfy the long-RT property, because they imply the condition $Q_{n-1}(t) = Q_n(t)$ for all $t > 0$. $\square$

An example of a (nonstandard) unlimited capacity model that does not satisfy the long-RT property is the model we considered earlier in which $Q_{n-1}(t) < Q_n(t)$ for $t < t'$ but $Q_{n-1}(t) > Q_n(t)$ for $t > t'$.

Again the result seems natural. In limited capacity models there is a decrease in processing rate when the load increases. Thus, overall, processing time should be lengthened as the processing load is increased. Alternatively, in supercapacity models processing time will tend to decrease with greater

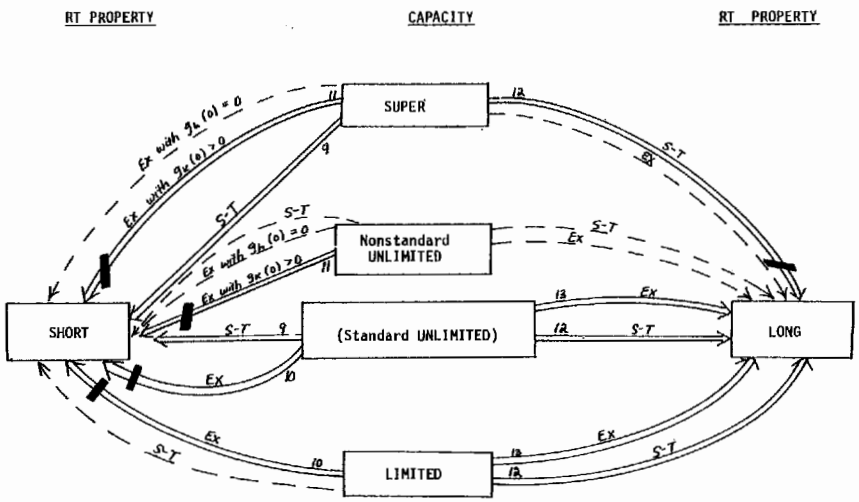


Fig. 8.2. A schematic summarizing the results of Propositions 8.9 through 8.13 (see text for explanation).

loads. This last result says that if this decrease is large enough, the long-RT property will be violated, even with a self-terminating search strategy.

Our intuitions about the long-RT property and limited capacity models turn out to be just as valid when search is exhaustive, as our next result shows.

Proposition 8.13: All limited capacity and all standard unlimited capacity, independent parallel exhaustive models predict the long-RT property to be true.

Proof: Limited capacity, independent parallel exhaustive models are those for which

$$\left[\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/(n-1)} \geqslant \left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{1/n} \quad \text{for all } t>0$$

Whenever this inequality holds, it is also true that

$$\prod_{i=1}^{n-1} G_{i,n-1}(t) \geqslant \left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{(n-1)/n} \geqslant \prod_{i=1}^{n} G_{i,n}(t) \quad \text{for all } t>0 \quad (8.2)$$

and thus that the long-RT property holds, $Q_{n-1}(t) \geqslant Q_n(t)$ for all $t>0$. For standard unlimited capacity models the leftmost inequality in Eq. 8.2 is a strict equality, and thus these models also predict the long-RT property to be true. □

Figure 8.2 schematizes the results of Propositions 8.9 through 8.13. The

Table 8.1. *Possible outcomes when applying the long- and short-RT properties and the classes of models whose members* all *predict those outcomes*

Long- and short-RT properties both hold for all $t>0$ Predicted by all *self-terminating* search models of *standard unlimited capacity*
Long-RT property holds for all $t>0$, *but short-RT property fails at some* $t'>0$ Predicted by all *exhaustive* search models with *standard unlimited capacity* or *limited capacity*
Long-RT property fails at some $t'>0$, *but short-RT property holds for all* $t>0$ Predicted by all *self-terminating* search models with *supercapacity*
Both properties fail at the same time $t'>0$ Conclusion: logically *impossible*

specified class predict the property. (2) the sign $\not\Rightarrow$ indicates that no models of the specified class can predict the property; equivalently, all models predict that the property is violated. (3) The sign $\dashrightarrow$ indicates that some of the models of the specified class predict the property whereas others violate it. It should also be recalled that the arrows involving $g_n(0)>0$ implicate only those models not possessing serial position effects.

Examining Fig. 8.2, we see that, as with serial models, the two RT properties do not logically discriminate between exhaustive and self-terminating search strategies in independent parallel models. The short-RT property, however, does a reasonably good job at this task, probably as well as it does with the serial models, although in connection with our discussion of Proposition 8.11 we saw that some independent parallel exhaustive models are able to predict the short-RT property.

The outcome is not so fortunate with the long-RT property. Large classes of exhaustive models predict the long-RT property to be true. Again this resembles the serial situation. With parallel processing the long-RT property appears to be more sensitive to the capacity issue than to differences in search strategy. Indeed, as Proposition 8.13 indicates, a long-RT property violation tends to point toward supercapacity, whereas verification of the property supports limited or standard unlimited capacity.

We have now investigated fairly thoroughly the question of which independent parallel models can and cannot predict the two RT properties. In most experimental applications, however, we are interested in answering a slightly different question. Suppose we have collected a set of RTs in some experimental setting and that we have tested these data for the long- and short-RT properties. The question is this: Given some specific outcome to these tests (e.g., long-RT property holds but short-RT property fails), what details can we infer about processing within the class of independent parallel models?

Table 8.1 attempts a partial answer to this question by conjoining the

cised in this venture because, it will be recalled, the capacity designations are somewhat different in the case of self-terminating as opposed to exhaustive processing. The arithmetic mean of the distribution functions was used to define capacity in self-terminating processing, whereas the geometric mean was used in the case of exhaustive processing. As noted earlier, these will be identical where serial position effects are nonexistent [i.e., when $G_{i,n}(t) = G_{j,n}(t)$ for $i \neq j$]. Although the exact definitions are somewhat fuzzy around the edges, no problem arises in the formulation of Table 8.1, since all the models exhibited there are totally self-terminating or totally exhaustive, thus bypassing the definitional fuzziness.

Table 8.1 lists only those classes of models *all* of whose members predict the particular empirical results. For instance, consider the case where both the long- and short-RT properties hold simultaneously. This result is predicted by all independent parallel self-terminating models with standard unlimited capacity, but Table 8.1 should not be interpreted as claiming that *no* other independent parallel models can predict the result. For instance, *some* limited capacity self-terminating models and some supercapacity exhaustive models can predict that both properties will hold (see Fig. 8.2). It may also be (and seems likely) that certain nonstandard unlimited capacity self-terminating models can simultaneously predict the satisfaction of the long- and short-RT properties, but this has not been shown.

The exact prediction boundaries of most other models involving the $\dashrightarrow$ are also not known at present. Therefore, Fig. 8.2 and Table 8.1 obviously exhibit models that are compatible with the long- and short-RT outcomes but do not determine all such models. "If and only if" statements between empirical results and model claims are preferable but often not feasible. An attempt to provide exact boundaries for some of these models is made in Tables 8.2 and 8.3. Each possible outcome imparts a set of bounds on the "average" individual item processing time distribution function (i.e., on

$$\frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t)$$

when search is self-terminating and on

$$\left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{1/n}$$

when search is exhaustive). These bounds, listed in the tables, allow us to make inferences about the capacity level of any system yielding these results.

Sometimes, rather than allowing us to specifically rule out a whole level of capacity, the bounds only allow us to make a weaker, more general statement. An example of this can be found in the first part of Table 8.3 in the case when both the long- and short-RT properties hold. Recall that when search is exhaustive, the long-RT property becomes

$$\prod_{i=1}^{n-1} G_{i,n-1}(t) \geq \prod_{i=1}^{n} G_{i,n}(t) \quad \text{for all } t>0$$

Table 8.2. *Conclusions that can be drawn about the capacity level of the processing system when the long- and short-RT properties are applied and search is self-terminating (see text for explanation)*

Long-RT and short-RT properties both hold

$$\frac{1}{n-1}\sum_{i=1}^{n-1} G_{i,n-1}(t) \geq \frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t) \geq \frac{1}{n}\sum_{i=1}^{n-1} G_{i,n}(t) \quad \text{for all } t>0$$

Conclusion: Capacity and model is unlimited or moderately limited.

Long-RT property holds and short-RT property fails at time t'

$$\frac{1}{n-1}\sum_{i=1}^{n-1} G_{i,n-1}(t) \geq \frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t) \quad \text{for all } t>0$$

and

$$\frac{1}{n}\sum_{i=1}^{n-1} G_{i,n-1}(t') > \frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t')$$

Conclusion: Capacity is unlimited or limited in general and is very limited at $t=t'$. Therefore, model is very limited in capacity.

Long-RT property fails at time t' *and short-RT property holds*

$$\frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t) \geq \frac{1}{n}\sum_{i=1}^{n-1} G_{i,n-1}(t) \quad \text{for all } t>0$$

and

$$\frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t') > \frac{1}{n-1}\sum_{i=1}^{n-1} G_{i,n-1}(t')$$

Conclusion: Overall capacity is not strongly limited (by top inequality) and it is supercapacity at time t' (bottom inequality). Therefore, model is supercapacity.

Both properties fail at time t'

$$\frac{1}{n-1}\sum_{i=1}^{n-1} G_{i,n-1}(t') < \frac{1}{n}\sum_{i=1}^{n} G_{i,n}(t') < \frac{1}{n}\sum_{i=1}^{n-1} G_{i,n-1}(t')$$

Conclusion: impossible.

which implies

$$\left[\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/n} \geq \left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{1/n} \quad \text{for all } t>0$$

but unfortunately does not imply

$$\left[\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/(n-1)} \geq \left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{1/n} \quad \text{for all } t>0$$

which is the definition of limited capacity. The bound does, however, place an upper limit, in terms of the $G_{i,n-1}(t)$, on how large

Table 8.3. *Conclusions that can be drawn about the capacity level of the system when the long- and short-RT properties are applied and search is exhaustive (see text for explanation)*

Long-RT and short-RT properties both hold

$$\left[\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/n} \geqslant \left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{1/n} \geqslant \left[\frac{n-1}{n}\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/n} \quad \text{for all } t>0$$

Conclusion: Capacity is super (by right-hand side) but not too super (by left-hand side). Therefore, model is moderately super. In addition, $g_n(0)=0$ by Proposition 8.11 (see text).

Long-RT property holds and short-RT property fails at time t′

$$\left[\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/n} \geqslant \left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{1/n} \quad \text{for all } t>0$$

and

$$\left[\frac{n-1}{n}\prod_{i=1}^{n-1} G_{i,n-1}(t')\right]^{1/n} > \left[\prod_{i=1}^{n} G_{i,n}(t')\right]^{1/n}$$

Conclusion: Capacity is not strongly super, and is even less so at time t'. Model could be mildly super, unlimited, or limited in capacity.

Long-RT property fails at time t′ *and short-RT property holds*

$$\left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{1/n} \geqslant \left[\frac{n-1}{n}\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/n} \geqslant \left[\frac{n-1}{n}\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/(n-1)} \quad \text{for all } t>0$$

and

$$\left[\prod_{i=1}^{n-1} G_{i,n-1}(t')\right]^{1/(n-1)} < \left[\prod_{i=1}^{n} G_{i,n}(t')\right]^{1/(n-1)} < \left[\prod_{i=1}^{n} G_{i,n}(t')\right]^{1/n}$$

Conclusion: Capacity is not severely limited (top inequality) and is supercapacity at time t' (bottom inequality). Therefore, capacity is super. Also, $g_n(0)=0$ by Proposition 8.11 (see text).

Both properties fail at time t′

$$\prod_{i=1}^{n-1} G_{i,n-1}(t') < \prod_{i=1}^{n} G_{i,n}(t') < \frac{n-1}{n}\prod_{i=1}^{n-1} G_{i,n-1}(t')$$

Conclusion: impossible

$$\left[\prod_{i=1}^{n} G_{i,n}(t)\right]^{1/n}$$

can be. Thus, although the upper bound is moot with respect to whether capacity is limited, it does tell us that it cannot be too strongly super. On the other hand, the lower bound reveals that capacity must be super, because

when t is small,

$$\left[\frac{n-1}{n}\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/n} > \left[\prod_{i=1}^{n-1} G_{i,n-1}(t)\right]^{1/(n-1)}$$

which implies supercapacity. For large t this inequality is reversed so the right-hand bound in Table 8.3 is weaker when t is large. However, since the Table 8.3 inequalities must hold for all t, supercapacity models are called for.

Figure 8.2 and Tables 8.1, 8.2, and 8.3 are, of course, logically compatible. It can be seen that Tables 8.2 and 8.3 permit model implications that are represented by dotted lines in Fig. 8.2 and thus are not representative of all models of that type. Put differently, Table 8.1 limits more severely the potential models because there it was demanded that all models of the class make a certain prediction.

Nevertheless, the three tables yield highly similar conclusions regarding capacity. The most diverse cases are found when the long- and short-RT properties both hold. In that instance, capacity must be super in the exhaustive models but unlimited or limited in the self-terminating models. Nevertheless, even in these cases, the exhaustive models cannot be very super, so the two properties establish firm capacity bounds for quite a broad class of parallel models.

Scrutiny of these three tables also indicates that if we consider supercapacity systems to be unlikely to be found in nature, then (1) "long- and short-RT properties hold" implies self-termination and unlimited or moderately limited capacity; (2) "long-RT property holds but short-RT property does not" implies unlimited or limited to very limited capacity, but processing could be self-terminating or exhaustive. The third possibility, that the long-RT property fails but the short-RT property holds, should almost never be found in data. Note that Sternberg's conclusion that case 1 implies self-termination follows in the wake of the ruling out of supercapacity, even though these models are not based on his other assumptions. The caveat should again be issued that these and the other conclusions of Fig. 8.2 and Tables 8.1, 8.2, and 8.3 depend on all the attendant assumptions mentioned earlier.

Empirical testing and tests

We are now in a position to discuss empirical applications of some of the nonparametric properties of processing time distributions we have been discussing. As we saw earlier, the assumptions of nested states (NS), accessibility ordering (AO), and tight packing (TP) lead to order properties on the average self-terminating processing distributions for those cases when a single target or critical element is embedded in an array of nontarget elements. This is the situation, of course, on target-present trials of both ET (early target or visual search tasks; e.g., Atkinson et al. 1969) and LT (late

target or memory-scanning tasks; e.g., Sternberg 1966) tasks. However, before we begin testing properties on these types of data, there is a very important question we must consider: What shall we take as evidence of the failure of, say, the long-RT property?

Empirical RT distribution functions are apt to be noisy under the best of conditions. If we take the null hypothesis to be

$$H_0: Q_{n-1}(t) \geq Q_n(t)$$

then a visual examination of the distribution functions with the decision rule of rejecting H_0 whenever there exists a single $t=t_0$ such that $Q_n(t_0) > Q_{n-1}(t_0)$ will probably result in a very large likelihood of a Type I error (falsely rejecting H_0). A statistical test is obviously called for.

One serious problem is that the natural direction to turn, the classic Smirnov tests (e.g., Walsh 1965), are based on the null hypothesis that the two distribution functions are identical, that is, that $Q_n(t)=Q_{n-1}(t)$. One can, however, utilize a one-sided Smirnov test with a null hypothesis of

$$H_0: Q_{n-1}(t) \geq Q_n(t) \quad \text{for all } t>0$$

The null hypothesis will be rejected only if there exists some t_0 for which $Q_n(t)$ is *significantly* greater than $Q_{n-1}(t_0)$. Note that the true probability of a type I error will typically be smaller than the chosen α since the latter is based on $Q_n(t)=Q_{n-1}(t)$, whereas it may be, as the model presumes, that $Q_{n-1}(t)>Q_n(t)$.

On the other hand, the Smirnov test, as is, is useless in detecting violations of the short-RT property since $nQ_n(t)$ and $(n-1)Q_{n-1}(t)$ have different asymptotes.

David S. Moore (personal communication, reported in Townsend 1974a) suggested a nonparametric test of the entire prediction that

$$\frac{n}{n-1} Q_n(t) \geq Q_{n-1}(t) \geq Q_n(t) \quad \text{for all } t>0$$

Let C_α be the upper critical point of the Smirnov test with significance level α. It is then proposed that H_0 (this ordering) be accepted if

$$Q_{n-1}(t)+2C_\alpha \geq Q_n(t)$$

and

$$\frac{n}{n-1} Q_n(t) + \frac{2n-1}{n-1} C_\alpha \geq Q_{n-1}(t) \quad \text{for all } t>0$$

and rejected otherwise. The first part just tests the long-RT property, but the second tests the short-RT property. It can be shown that the probability that the predicted order is observed, given it is true, is greater than or equal to $(1-\alpha)^2$. That is, the true probability of a Type I error is less than or equal to that chosen for the regular Smirnov test. Unfortunately, the bounds set by this test appear to be so large that although the true probability of incorrectly

rejecting H_0 is very small, it appears that the test lacks power; that is, the probability of incorrectly accepting H_0 appears to be quite large, given reasonable violations of one or both of the two RT properties. Our explorations so far indicate that this problem is real and so the test will not be used below. Nevertheless, it may be feasible in the future to alter it in some fashion that will render it more practical.

The best strategy at present seems to be to reject the null hypothesis of the short-RT property,

$$H_0: nQ_n(t) \geq (n-1)Q_{n-1}(t) \quad \text{for all } t>0$$

whenever there is any $t=t_0$ such that $nQ_n(t_0) < (n-1)Q_{n-1}(t_0)$, even though the likelihood of a Type I error will no doubt be large. Thus it must be kept in mind throughout the following discussion that whereas a fairly good statistical test exists for the long-RT property, no such test has been constructed for the short-RT property. Nevertheless, it will be seen below that the violations of the short-RT property tend (so far) to be so large and frequent as to render this problem academic.

Applications

Sternberg (1973) reports several empirical tests of his RT properties. The first was a choice RT task with a 1:1 stimulus–response mapping. The stimuli were the numerals 1 through 8, and the responses were key presses. The different conditions involved differing numbers of S-R pairs (2, 4, and 8). Mean RT was found to increase linearly with the logarithm of the number of S-R pairs. For all observers there were large violations of the short-RT property but none of the long-RT property. From our earlier investigations we know this pattern of results suggests a system quite limited in capacity and in which one or more of TP, NS, and self-termination fails.

The second experiment was also a choice RT task with the numerals 1 through 8 as stimuli. Here, however, responses were vocalizations of the numeral names. The conditions involved 2 and 8 S-R pairs. All observers again violated the short-RT property, but this time 4 of the 5 also violated the long-RT property. It will be recalled that a failure of the long-RT property cannot be predicted by any self-terminating or exhaustive model satisfying TP, NS, and AO. However, parallel supercapacity models do predict this violation, and some of these models will also predict a short-RT property violation if an exhaustive search is assumed (see Proposition 8.11 and Fig. 8.2). It is difficult to imagine an experimental task more amenable to a supercapacity type of processing than the naming of arabic numerals. This explanation is further supported in comparisons with the key press experiment. There, mean RT increased roughly 150 msec between the 2-choice and the 8-choice conditions. When the response was changed from depressing keys to naming the stimuli, the increase between the 2-choice and 8-choice conditions fell to only about 20 msec.

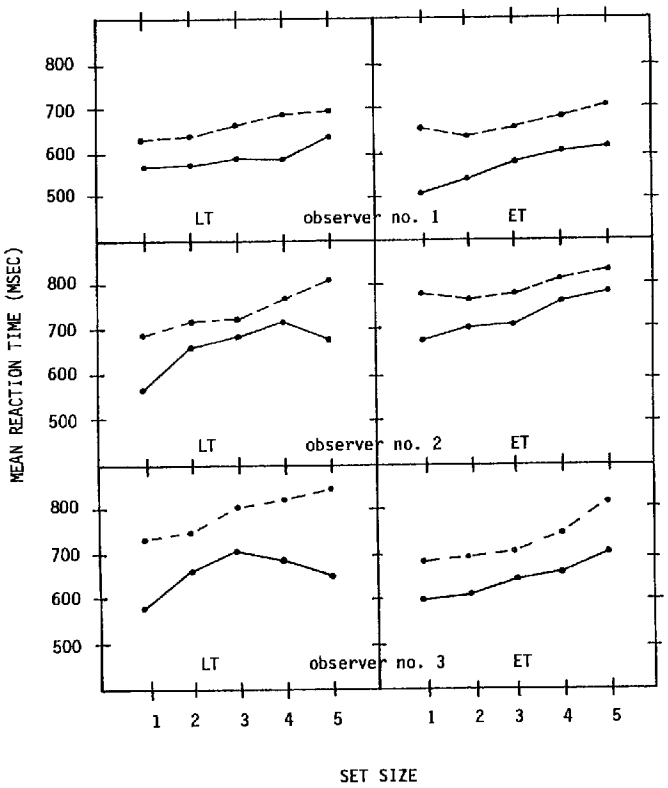


Fig. 8.3. Target-present and target-absent mean RT curves for the three observers in the experiment of Townsend and Roos (1973). The left-hand panels are from a memory scanning (LT) task and the right-hand panels summarize the results of a visual search (ET) task.

In an effort to extend the applications of these two properties we applied them to the binary classification study of Townsend and Roos (1973).[2] Three observers participated in both an LT task (late target or memory-scanning task; e.g., Sternberg 1966) and an ET task (early target or visual search task; e.g., Atkinson et al., 1969). The stimulus lists contained from one to five English consonants, and in all cases the target displays contained exactly one English consonant. One-half of the trials were *target-present* trials in which the target consonant was contained in the list; the other half were *target-absent* trials.

In general, mean RT tended to increase linearly with the size of the search set in both conditions (see Fig. 8.3). Target-present and target-absent curves tended to be parallel, with slopes ranging from 15 to 25 msec depending upon the observer and the condition. Error rates were approximately 1.8% in both

conditions. These trials and all those on which RT was less than 100 msec or greater than 2,000 msec (about 1%) were excluded from analysis.

The long-RT property was tested in a two-step procedure. First a two-sided Smirnov test was employed. When H_0 was accepted in this test, it was concluded that $Q_n(t)=Q_{n-1}(t)$ for all $t>0$, and thus the long-RT property was verified at the degenerate level. Note that if $Q_n(t)=Q_{n-1}(t)$ for $n=1$ to 4, then capacity is unlimited at the level of the total completion time since the RT distribution is unaffected by an increase in the load. The second step in the test involved a one-sided test with H_0: $Q_{n-1}(t) \geqslant Q_n(t)$ for all $t>0$, and H_1: $Q_{n-1}(t)<Q_n(t)$ for some $t>0$.

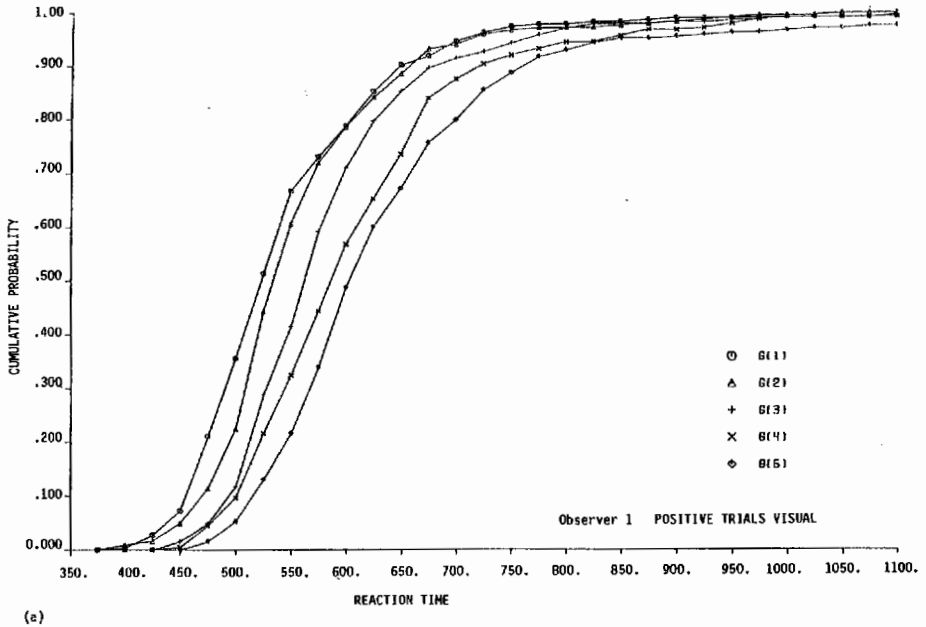
If H_0 was rejected in the two-sided test and accepted in the one-sided test, then it can be concluded that $Q_{n-1}(t)>Q_n(t)$. If H_0 was accepted in the two-sided test, then it was concluded that $Q_n(t)=Q_{n-1}(t)$. If H_0 was rejected in the two-sided test and rejected in the one-sided test, it follows that some RTs exist such that $Q_{n-1}(t)<Q_n(t)$ and hence the long-RT property is falsified.

These tests were performed on all pertinent distribution functions for the target-present as well as the target-absent trials. The latter should, of course, continue to satisfy the long-RT property even though processing must be exhaustive with respect to the letter matches. Out of the 120 tests, 22 accepted H_0 in the degenerate form $Q_n(t)=Q_{n-1}(t)$ with $\alpha=.05$ and 96 accepted H_0 in the form $Q_{n-1}(t)>Q_n(t)$. Thus, in only 2 instances was the long-RT property rejected (observer 1, target absent, ET, $n=1$ and observer 2, target present, LT, $n=4$). In both cases, there was a concomitant decrease in the mean RTs. Since the ordering $Q_{n-1}(t) \geqslant Q_n(t)$ at the individual observer level implies the correct ordering when the data is pooled, we show in Fig. 8.4a a typical observer and then in Fig. 8.4b the grouped data for the target-present trials in the ET condition. Note that there is a crossover in the pooled data of $Q_5(t)$ for short RTs and of $Q_2(t)$ at the long RTs. Observer 2 evidenced the first crossover, whereas observers 2 and 3 showed the latter. These differences were statistically nonsignificant and in any case occupy only the extremes of the RT range.

Our investigation of the short-RT property must be somewhat tentative because (a) the property only applies to the left tail of the distributions, since it holds a priori as t becomes large; (b) in connection with (a) the exact range where the short-RT property should hold will not typically be known; (c) being confined to the tail, the test will be more open to sampling error and anticipation RTs; and (d) no adequate statistical tests of this property are now known. With these caveats in mind we looked over the data and found many apparent violations of the short-RT property.

Figures 8.5a and b respectively show estimates of $nQ_n(t)$ for $n=1$ to 5 for an individual observer (same as Fig. 8.4a) and for all observers grouped together for target-present trials in the ET condition. Sizable violations seem to be present here as in the other reported tests on the short-RT property.

These data seem to indicate that (a) we are dealing with systems that are basically limited capacity; hence the acceptance of the long-RT property; but



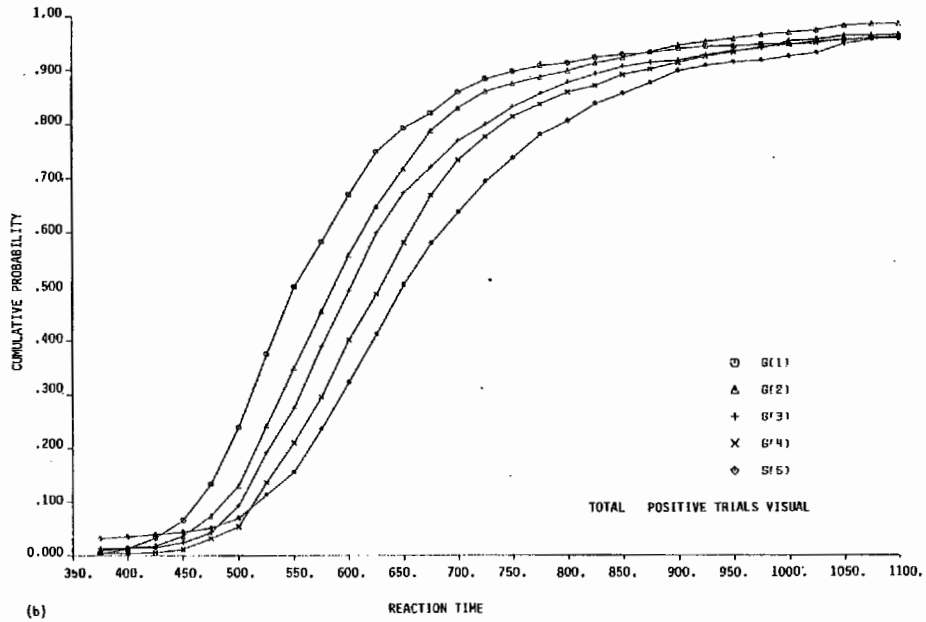
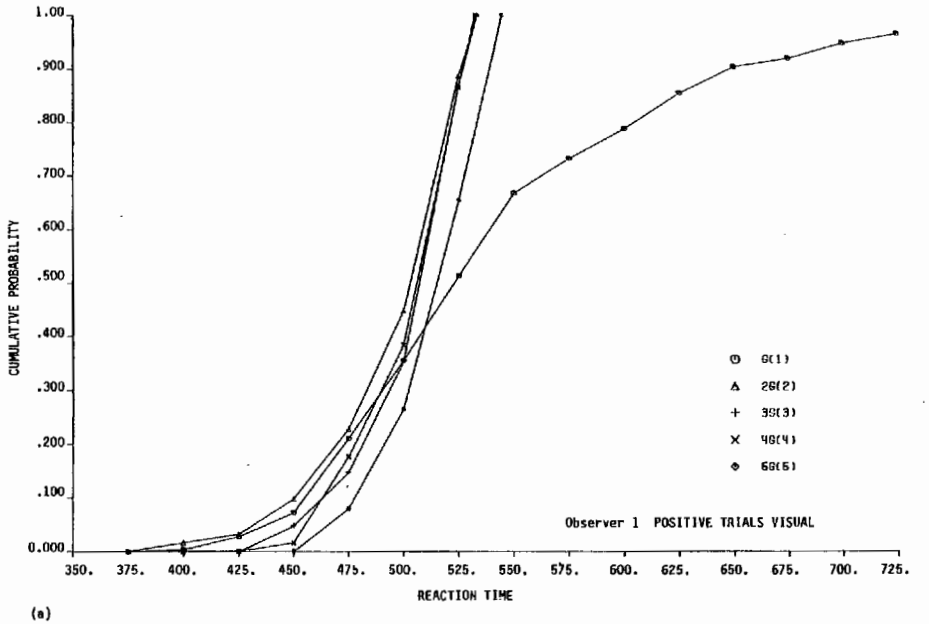


Fig. 8.4. Examples of applying the long-RT property to the target-present data from the visual search (ET) task of Townsend and Roos (1973). (a) shows the estimated RT distribution functions, $Q_n(t)$, of observer 1 and (b) shows the $Q_n(t)$ estimates when the data of the three observers are grouped together. The long-RT property predicts the ordering, $Q_1(t) \geqslant Q_2(t) \geqslant \cdots \geqslant Q_5(t)$ for all $t > 0$.



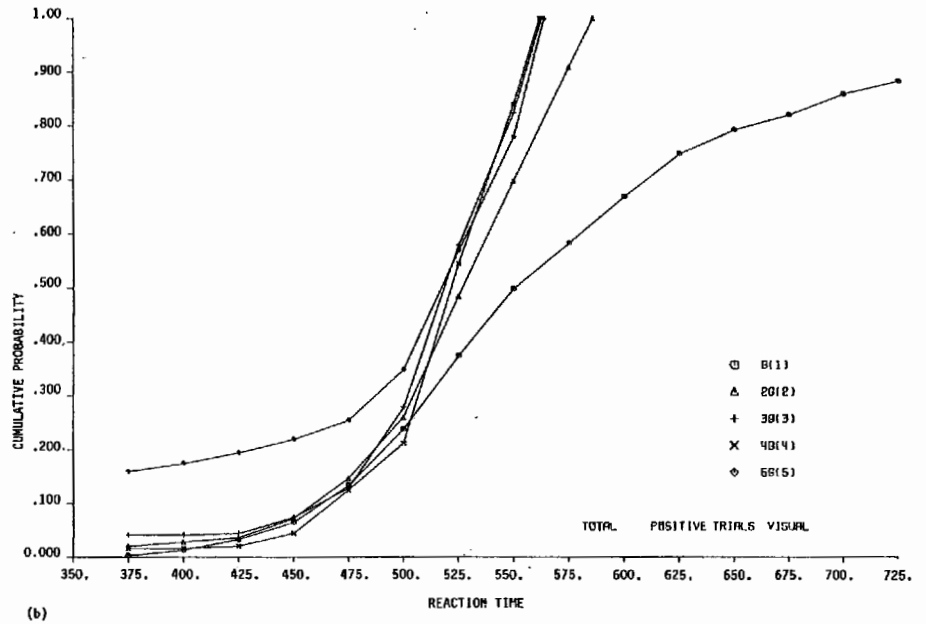


Fig. 8.5. Examples of applying the short-RT property to the target-present data from the visual search (ET) task of Townsend and Roos (1973). (a) shows the estimate of the function, $nQ_n(t)$ for observer 1 and (b) shows the $nQ_n(t)$ estimates when the data of the three observers are grouped together. The short-RT property predicts these functions should be ordered as $5Q_5(t) \geqslant 4Q_4(t) \geqslant \cdots \geqslant Q_1(t)$ for all $t > 0$. Since this ordering is guaranteed for large t, only small t are plotted in the figure.

[i.e., we occasionally find $Q_n(t)=Q_{n-1}(t)$]; and (c) the capacity limitations as n increases are often too severe to allow satisfaction of the short-RT property. A glance back to Tables 8.2 and 8.3 tells us that these conclusions are valid in the case of either a self-terminating or an exhaustive search.

We may amalgamate the present findings with other aspects of the data to reach some general tentative conclusions. First, the presence of strong serial position mean RT effects (see Townsend & Roos 1973) also shows up in the RT distribution orderings on target-present trials and thus must be considered a powerful phenomenon. Second are the roughly parallel target-present and target-absent mean RT curves of Fig. 8.3. Both mean RT serial position effects and parallel target-present and target-absent mean RT curves have been reported in many studies. Amalgamating the in-depth discussion of such topics in Chapter 7, we believe a limited-capacity parallel self-terminating model is most compatible with the overall pattern of results in this and a number of comparable studies.

Capacity during the minimum completion time

Before closing this chapter we shall take time to examine one more level of processing that has previously been touched on only lightly. This is the minimum completion time or, equivalently, the first intercompletion time or the first stage of processing. If search is serial, the first intercompletion time is just the individual item processing time of the first item completed. However, when search is parallel, all processors are active during the first stage, and as we shall soon see, an examination of the capacity during the first stage of processing when search is parallel and within-stage independent or completely independent proves to be especially illuminating. For this reason, the discussion of this last section will focus on parallel models.

We could adopt the strategy, which we have followed throughout this chapter, of examining the cumulative distribution function for the first stage of processing. However, it turns out there is another level at which we can describe capacity that is more sensitive than the distribution function and is much better suited to describing events during the first stage of processing. This is the so-called hazard function (see Chapter 3) defined by

$$H(t)=\frac{g(t)}{1-G(t)}$$

The hazard function is more sensitive than the distribution function in the sense that an ordering on hazard functions implies a distribution function ordering, but not vice versa (Townsend & Ashby 1978). Thus $H_A(t)\geqslant H_B(t)$ for all $t>0$ implies $F_A(t)\geqslant F_B(t)$ for all $t>0$; but if we know the distribution functions are ordered, we cannot be certain the hazard functions are.

Since all processors are active during the first stage of processing when search is parallel, the overall first-stage hazard function will be a combination of the hazard functions of all the elements that are to be processed. As

the next result shows, this combination turns out to be extremely simple if processing is within-stage-independent, because this type of independence guarantees that any change in the individual element hazard function for an element on one channel does not affect the hazard functions associated with any of the other channels.

Proposition 8.14: If processing is parallel and within-stage independent, then the hazard function $H_n(t)$ of the first intercompletion time when there are n elements to be processed is equal to the sum of the individual element hazard functions $H_{i,n}(t)$; that is,

$$H_n(t)=\sum_{i=1}^{n} H_{i,n}(t)$$

Proof: With one element, the first-stage hazard function is, of course, just

$$H_1(t)=\frac{g_{1,1}(t)}{1-G_{1,1}(t)}=\frac{g_{1,1}(t)}{\bar{G}_{1,1}(t)}$$

With two elements in the display, the probability density that the element in position 1 is completed first is

$$g_{1,2}(t)[1-G_{2,2}(t)]=g_{1,2}(t)\bar{G}_{2,2}(t)$$

The complete first-stage density function must also account for those instances when the item in position 2 finishes first; thus

$$g_2(t)=g_{1,2}(t)\bar{G}_{2,2}(t)+g_{2,2}\bar{G}_{1,2}(t) \tag{8.3}$$

The corresponding distribution function is

$$G_2(t)=1-\bar{G}_{1,2}(t)\bar{G}_{2,2}(t)=G_{1,2}(t)+G_{2,2}(t)-G_{1,2}(t)G_{2,2}(t)$$

Note that differentiation of $G_2(t)$ yields

$$\begin{aligned}g_2(t)&=g_{1,2}(t)+g_{2,2}(t)-g_{1,2}(t)G_{2,2}(t)-G_{1,2}(t)g_{2,2}(t)\\&=g_{1,2}(t)[1-G_{2,2}(t)]+g_{2,2}(t)[1-G_{1,2}(t)]\end{aligned}$$

which agrees with Eq. 8.3. Finally we can write the survivor function as

$$\begin{aligned}\bar{G}_2(t)&=1-G_2(t)=1-G_{1,2}(t)-G_{2,2}(t)+G_{1,2}(t)G_{2,2}(t)\\&=[1-G_{1,2}(t)][1-G_{2,2}(t)]=\bar{G}_{1,2}(t)\bar{G}_{2,2}(t)\end{aligned}$$

We now have all of the necessary quantities to write the two-element first-stage hazard function:

$$\begin{aligned}H_2(t)&=\frac{g_2(t)}{\bar{G}_2(t)}=\frac{g_{1,2}(t)\bar{G}_{2,2}(t)+g_{2,2}(t)\bar{G}_{1,2}(t)}{\bar{G}_{1,2}(t)\bar{G}_{2,2}(t)}\\&=\frac{g_{1,2}(t)}{\bar{G}_{1,2}(t)}+\frac{g_{2,2}(t)}{\bar{G}_{2,2}(t)}=H_{1,2}(t)+H_{2,2}(t)\end{aligned}$$

This development generalizes to an arbitrary number of elements in a straightforward manner. With a load of n elements,

$$H_n(t)=\frac{g_n(t)}{\bar{G}_n(t)}=\frac{\sum_{i=1}^{n} g_{i,n}(t)\prod_{\substack{j=1\\ \neq i}}^{n}\bar{G}_{j,n}(t)}{\prod_{i=1}^{n}\bar{G}_{i,n}(t)}=\sum_{i=1}^{n}\frac{g_{i,n}(t)}{\bar{G}_{i,n}(t)}$$

$$=\sum_{i=1}^{n}H_{i,n}(t)\ \square$$

Thus, the overall first-stage hazard function can be written as the sum of the individual element hazard functions. In Chapter 4 we identified the hazard function with the capacity of a system and thus Proposition 8.14 states the very appealing result: To calculate the first-stage capacity we need only sum the capacities allocated to each of the individual elements.

Unlimited capacity, independent parallel models

The above convenient property, that the first-stage hazard function equals the sum of the individual element hazard functions, allows very straightforward definitions of our three types of capacity at this level of processing (Townsend 1974b). A system or model is called *unlimited capacity* on the *first* stage of processing if $H_n(t)=nH_1(t)$. In other words, in an unlimited capacity system the sum of the individual element hazard functions, when there are a total of n elements, is equal to n times the hazard function when there is only one element; adding more elements to the display does not affect the individual element hazard functions.

We can reduce this expression to one on the survivor functions because this definition of unlimited capacity implies that

$$\sum_{i=1}^{n}\left[\frac{g_{i,n}(t)}{\bar{G}_{i,n}(t)}\right]=n\left[\frac{g_{1,1}(t)}{\bar{G}_{1,1}(t)}\right]$$

Integration of both sides yields

$$\int_0^t\sum_{i=1}^{n}\left[\frac{g_{i,n}(s)}{\bar{G}_{i,n}(s)}\right]ds=\sum_{i=1}^{n}\int_0^t\left[\frac{g_{i,n}(s)}{\bar{G}_{i,n}(s)}\right]ds=\int_0^t n\left[\frac{g_{1,1}(s)}{\bar{G}_{1,1}(s)}\right]ds$$

Because $g_{i,n}(t)$ equals -1 times the derivative of $\bar{G}_{i,n}(t)$ these integrations are easily performed, yielding

$$-\sum_{i=1}^{n}\log\bar{G}_{i,n}(t)=-n\log\bar{G}_{1,1}(t)$$

This expression can be simplified to

$$\log\left[\prod_{i=1}^{n}\bar{G}_{i,n}(t)\right]=\log\left[\bar{G}_{1,1}(t)\right]^n$$

Therefore, in independent parallel models unlimited capacity during the first stage of processing implies

$$\prod_{i=1}^{n}\bar{G}_{i,n}(t)=[\bar{G}_{1,1}(t)]^n$$

or equivalently

$$\left[\prod_{i=1}^{n}\bar{G}_{i,n}(t)\right]^{1/n}=\bar{G}_{1,1}(t) \tag{8.4}$$

This hazard function definition of capacity therefore imposes natural constraints on the geometric means of the individual item *survivor* functions, and it was this result that motivated Definition 8.3. The use of survivor functions is in contrast to our capacity definitions based on the level of the individual item completion times that we formulated directly on the geometric means of the *distribution* functions in the case of exhaustive processing (Definition 8.2). Of course, a constraint imposed on survivor functions implies a concomitant constraint on the distribution functions, but unfortunately no simple mapping between the geometric means of survivor functions and distribution functions exists. For this reason it is difficult to analytically compare the two definitions.

The Eq. 8.4 result seems very natural: The (geometric) average of the $\bar{G}_{i,n}(t)$ equals the survivor function when $n=1$; in other words, the probability of not having completed each individual element when the load is n elements is on the average equal to the probability of not having completed one item when the load is one element.

This same idea is readily seen in the hazard function definition, $H_n(t)=nH_1(t)$ when we divide through by n,

$$\frac{1}{n}H_n(t)=H_1(t)$$

In unlimited capacity models, the average capacity allocated to each of the n elements is equal to the capacity allocated to one element when it is the only one being processed.

Limited capacity, independent parallel models

An independent parallel model is called *limited capacity* on the first stage of processing if $H_n(t)<nH_1(t)$. This implies, of course, that

$$\frac{1}{n}H_n(t)<H_1(t)$$

That is to say, the average capacity assigned to each of the n elements is less than the capacity assigned to one element when it alone is being processed.

In a development analogous to the one when capacity is unlimited, it can

be shown that the above definition of limited capacity implies the following order on the survivor functions:

$$\prod_{i=1}^{n} \bar{G}_{i,n}(t) > [\bar{G}_{1,1}(t)]^n$$

or equivalently,

$$\left[\prod_{i=1}^{n} \bar{G}_{i,n}(t)\right]^{1/n} > \bar{G}_{1,1}(t) \tag{8.5}$$

Equation 8.5 states that the (geometric) average probability that all of the n elements are still undergoing processing is greater than the probability that a single element is still being processed when it constitutes the entire load on the system.

This definition of limited capacity does not mean that there is necessarily only a *fixed* amount of capacity available. The fixed capacity model is a special case of limited capacity in which $H_n(t) = H_1(t)$. Here it is as if, no matter what the processing load, the system always has available the same amount of capacity that it must, in some way, spread out over the n element positions. This means that the sum of the n individual element capacities must be a constant value, independent of n.

It should be emphasized that this class of models does not necessarily imply a separate fixed source of capacity, although that is one reasonable realization of the model. Any number of systems could give results of this kind. For instance, a system with a moderate but not a high degree of mutual channel interference (e.g., lateral interference in visual paradigms) could yield fixed capacity predictions. Clearly, analogous remarks can be made for the other brands of systems as well (see Chapter 6).

Models that are even more limited in capacity than fixed capacity are called *extremely limited capacity* models. In these models, $H_n(t) < H_1(t)$.

Supercapacity, independent parallel models

A model with capacity that increases on each position or channel with each additional item to be processed is said to be of *supercapacity* at the level of the first stage of processing. We may write this in terms of the hazard functions as $H_n(t) > nH_1(t)$ or

$$\frac{1}{n} H_n(t) > H_1(t)$$

It is easily seen that this definition implies the following order on the corresponding survivor functions:

$$\prod_{i=1}^{n} \bar{G}_{i,n}(t) < [\bar{G}_{1,1}(t)]^n$$

As expected, the first completion time occurs stochastically earlier with n items than would be the case with n replications of the $n=1$ situation.

An ordered, limited capacity parallel model

A rather interesting model results when the following ordering is assumed among the first intercompletion time hazard functions as n is incremented:

$$H_{n-1}(t) \leqslant H_n(t) < \frac{n}{n-1} H_{n-1}(t)$$

It can be quickly ascertained that $H_1(t) \leqslant H_n(t)$, so the system represented by the model has capacity bounded below by the fixed capacity model. Also, note that $H_n(t) < nH_1(t)$, and so it is bounded above by the unlimited capacity model. Thus its capacity does not decrease as n increases but also does not increase in such a way that it becomes unlimited or supercapacity.

Notice that this same ordering could be rewritten as

$$\frac{1}{n} H_{n-1}(t) \leqslant \frac{1}{n} H_n(t) < \frac{1}{n-1} H_{n-1}(t)$$

and therefore it implies the following ordering on the survivor functions:

$$\left[\prod_{i=1}^{n-1} \bar{G}_{i,n-1}(t)\right]^{1/(n-1)} < \left[\prod_{i=1}^{n} \bar{G}_{i,n}(t)\right]^{1/n} \leqslant \left[\prod_{i=1}^{n-1} \bar{G}_{i,n-1}(t)\right]^{1/n}$$

For example, when n is 3, this ordering becomes

$$[\bar{G}_{1,2}(t)\bar{G}_{2,2}(t)]^{1/2} < [\bar{G}_{1,3}(t)\bar{G}_{2,3}(t)\bar{G}_{3,3}(t)]^{1/3} \leqslant [\bar{G}_{1,2}(t)\bar{G}_{2,2}(t)]^{1/3}$$

For brevity let us define

$$\pi_1 = \bar{G}_{1,1}(t)$$

$$\pi_2 = \bar{G}_{1,2}(t)\bar{G}_{2,2}(t)$$

$$\pi_3 = \bar{G}_{1,3}(t)\bar{G}_{2,3}(t)\bar{G}_{3,3}(t)$$

Then the above ordering becomes $\pi_1 < (\pi_2)^{1/2} \leqslant (\pi_1)^{1/2}$ when $n=2$ and $(\pi_2)^{1/2} < (\pi_3)^{1/3} \leqslant (\pi_2)^{1/3}$ when $n=3$. These relations are illustrated in Fig. 8.6. Note that although the figure shows $(\pi_1)^{1/2} \leqslant (\pi_3)^{1/3}$, this is not implied by the ordering relations.

Many of the ideas broached in this section can be generalized to higher stages of processing (see the section "Construction and Analysis of Models via Intensity Function Dynamics" in Townsend & Ashby 1978).

Hazard function definitions of capacity are associated with some very nice properties, specifically that the hazard function (i.e., the capacity) of the first intercompletion time is equal to the sum of the individual element hazard functions (i.e., capacities). Because of such properties, these definitions hold

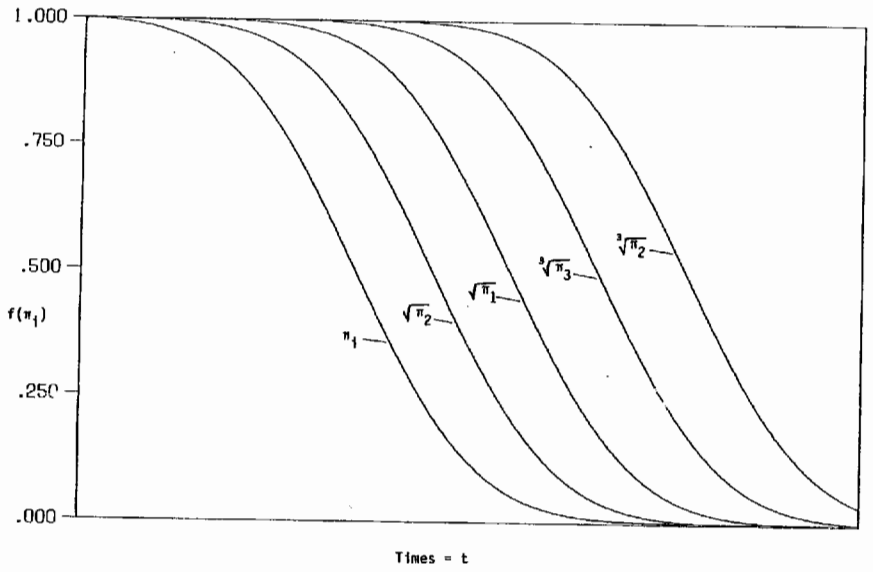
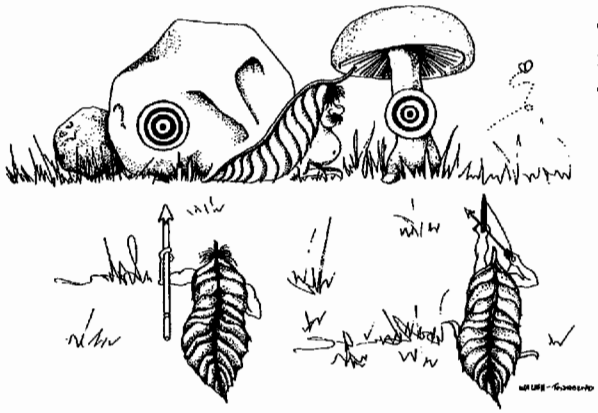


Fig. 8.6. Orderings on sets of survivor functions as implied by the ordered limited capacity parallel model described in the text.

great promise. The drawback is that hazard functions impose natural constraints on survivor functions and not distribution functions and at this point in time, distribution functions are the more popular statistics.

In the foregoing developments, we have seen how distributional properties, apart from the often-applied central moments, can form a promising arsenal for testing models and analyzing interesting system dynamics. Non-moment-distributional characteristics will play a lesser but still major role in the following chapter, which investigates varied-state and counting models.



These two Leafmen are illustrating the principle of *speed-accuracy trade-off*. The Leafman with the bow and arrow will shoot from a considerable distance at (of course) high speed, while the Leafman with the spear plans to run up to the target and plunge it into the center. Thus, we expect that the spear-toting Leafman will be *slow* but *accurate,* whereas the archer Leafman will be *fast* but *inaccurate.*

9 *Reaction time models and accuracy losses: varied state and counting models*

Many of the models we have discussed so far in this book assume that the observer's processing system operates in an error-free manner, or at least that the few errors that do occur do not fundamentally disrupt the nature of processing. This situation is of interest, of course, since error rates in many pattern-matching studies are very low. There is evidence that so long as the overall error rate is less than about 10% in some of the paradigms of interest in this book, the errors that do occur tend to come from the response selection stage rather than from comparison (Swanson & Briggs 1969; Pachella 1974).

Even so, there are many situations in which error rates are substantial. The human observer is not perfect, and an understanding of how and when errors occur is of fundamental importance to the understanding of the human processing system. In this and the next chapter we focus on situations with substantial accuracy losses, and we will investigate reaction time models imbued with enough structure to be capable of predicting both RT and accuracy.

We inspect three large classes of models: varied-state models, counting models, and random walk models. Although bearing some similarities, these large classes differ from each other in certain fundamental respects.

Varied-state models, which are related to many of the models we considered in the preceding chapter, assume that the experimental variability in

an observer's performance is the result of mixing several distinct types of response, each of which is associated with a different internal state. On any single experimental trial the observer is assumed to enter into a distinct internal state with some probability. Thus, there is assumed to be a probability distribution over a (usually small) set of internal states. Associated with each of these states is an error rate and a distribution of response times.

Perhaps the best known of these are the *fast-guess* models developed by Ollman (1966) and Yellott (1967, 1971). In the prototypical fast-guess model there are two internal states: a stimulus-controlled state and a guessing or no-information state. In the former, the observer is assumed to fully process the stimulus and hence accuracy and response time are both high. When the observer is in the guessing state, he or she is assumed to prematurely terminate processing of the stimulus and guess a response. On these trials, response times will be short, but a price will be paid in the form of decreased accuracy.

Varied-state models tend to be psychologically intuitive, but often at only a fairly macroscopic level. They are usually not amenable to a detailed (e.g., physiologically suggestive) processing interpretation, as are the counting models, for instance. The notion of internal states is generally meant to reflect the belief that observers can perform a task by choosing one of several higher-level cognitive processing strategies or perhaps that the observer may fluctuate between a stimulus and a response orientation. Another possibility in the case of the simpler fast-guess model is that, relative to a particular set of stimuli, observers may be processing in an all-or-none manner; that is, each stimulus is either perceived perfectly or no information at all is accrued (see Townsend 1971a, b).

Counting models and random walk models are both special cases of varied-state models in which there is only one internal "state," since they both assume that basically the same thing happens on every trial. Both classes assume the observer accumulates information about the stimulus over time. When the information in favor of any one response is strong enough, processing is halted and that response is made. The two classes differ in their notions of how this accumulation of information proceeds.

Counting models postulate the existence of a separate internal "counter" for each stimulus alternative. When a stimulus is presented, each counter begins "looking" for evidence favoring its alternative. It signals its presence when some preset criterion of evidence is exceeded. The response is determined by the first counter to emit a positive signal. Thus, processing becomes a race between counters. The first one to reach its criterion "wins," and its alternative is chosen as the one most likely to have been presented on that trial.

Although it is not necessary to do so (Vickers 1979), evidence is usually assumed to be accumulated in discrete chunks, and each chunk is assumed to be as informative as any other. Thus, the counter criteria are a set of integers,

each specifying the number of discrete chunks of evidence, or alternatively, the number of counts that must be accumulated on each counter before the presence of its stimulus alternative is signaled.

Compared to the other models we will study in this and the next chapter, counting models have a fairly long and rich history, going back at least as far as the early 1960s (LaBerge 1962). Pike, among others, studied counter models extensively (Audley & Pike 1965; Pike 1966, 1968, 1971, 1973). McGill is responsible for an unpublished (so far) work on counting models that has become something of an underground classic (McGill 1964). Much of the early work focused on the accuracy of counters and ignored the question of response latency. However, it is very natural to think of the process of accumulating counts as being time-consuming, so that the higher the count criteria, the slower the response time. All that is needed to generate RT predictions is an assumption about the time it takes to accumulate counts. We shall examine the latency predictions of counting models in detail in this chapter.

Although processing interpretations of counting models can be made at many different levels, there is sometimes a natural interpretation of counts as action potentials or neural spikes, which are propagated down some neural pathway. For instance, consider a visual recognition experiment where the stimuli are horizontal and vertical line segments. In modeling this task, we might suppose that, somewhere centrally, there are horizontal and vertical feature detectors that receive spikes at a maximal rate when a horizontal and vertical line segment are presented, respectively. The recognition process now becomes a matter of counting the spikes that are input to the two feature detectors and making the appropriate response when one of the two records some predetermined number.

A slight generalization of this are models that assume perceived stimulus intensity is proportional to the number of spikes received per unit time so that computing intensity is equivalent to counting spikes. Models of the human sensory system based upon the counting of neural spikes have their own fairly old and rich history (Barlow 1956, 1957; Bouman 1961; Creelman 1961; Kohfeld 1969; McGill 1963, 1967; McGill & Goldberg 1968; Siebert 1965, 1968, 1970). Another closely related possibility is to judge stimulus intensity by timing how long it takes some fixed number of spikes to arrive. The perceived intensity of the stimulus will increase inversely with the time it takes to count a fixed number. Neural timing models were extensively investigated by Luce and Green (1970, 1972; see also Green & Luce 1967, 1971, 1972).

Like counting models, random walk models assume that information is accumulated over time. The difference between the two is in the number of counters postulated. Random walk models assume that in a two-choice situation there is only one counter, but that it has two criteria. When evidence is found for one of the alternatives, say alternative S_A, the counter is incremented, and when evidence is found for the other alternative, S_B, it is decre-

mented. When the evidence count exceeds some positive criterion, alternative S_A is chosen, and when it falls below some negative criterion, alternative S_B is chosen.

Thus, in random walk models, evidence for one alternative is viewed not only as evidence in its favor but also as evidence *against* the other alternative. One way of generating a random walk model is to start with a counting model and then impose a differencing mechanism and finally a counter that counts these differences. The criterion is therefore relative, rather than absolute, as in counting models. The model derives its name because processing can be viewed as a random walk between the positive and negative criteria. Although the count will tend to drift toward the correct criterion, there will still be randomness in its moment-by-moment location. We will examine the predictions of random walk models in the next chapter.

An experimental overview

Before we begin examining models in detail it will help to have a broad overview of the types of experimental results that we expect any candidate model to be able to predict. As a representative task we shall consider a two-alternative forced choice detection paradigm similar to the one developed by Estes and Taylor (1964). An example of such a task is the lateral masking study of Townsend and Snodgrass (1977).

In the latter study, two sets of target letters were used; the capital letters D and O formed one set, and I and L formed the other. One of these groups served as the target set for each block of trials. On every trial a masking letter appeared to the right or the left of the target letter. One of these letters, D, O, I, or L, was selected randomly to serve as the mask, regardless of which group was the target set for that block of trials. The observer's task was to respond as quickly as possible (with a button press) depending on the identity of the presented target element.

The most striking result of the study was the failure to find uniform effects of the target–mask similarity on detection performance. For instance, I and L tended to mask the target O more effectively, but the masks O and D were more effective when the target was D. To a certain extent, it appears likely that at least some of these effects are specific to the particular target–mask stimulus lists employed. For example, one possible interpretation of the results is that when the target set is $\{O, D\}$, the observer adopts the criterion of responding D as soon as a vertical line feature is extracted from the display. If only curved features are extracted, the observer may respond O. Under this strategy an I or L mask might facilitate D responses because of the increased number of straight-line features.

In addition, and more pertinent to the present discussion, it was found that correct and error mean RTs were more dependent on the response than on the stimulus. This result is illustrated in Fig. 9.1, which gives the four mean RT combinations for each target set. Let $\bar{T}_{ij}$ be the mean RT for response j to

Stimulus \ Response	O	D
O	754	874
D	806	821

Stimulus \ Response	I	L
I	751	859
L	794	783

Fig. 9.1. Mean RT in msec (Townsend & Snodgrass 1977).

stimulus i. It can be seen that the orderings are in the relation $\bar{T}_{OO} < \bar{T}_{DO} < \bar{T}_{DD} < \bar{T}_{OD}$ and $\bar{T}_{II} < \bar{T}_{LL} < \bar{T}_{LI} < \bar{T}_{IL}$.

Note that for both sets of stimuli mean correct response latencies are faster than incorrect latencies (averaged over the two stimuli in a stimulus set). Faster correct latencies are very frequently found in the literature, particularly when the discrimination is difficult (Audley 1973; Estes & Wessel 1966; Pike 1968). However, when discriminations are easier it is not uncommon to find the reverse result that errors are faster than correct responses (Ollman 1966; Swensson & Edwards 1971; Yellot 1967). Indeed, this latter pattern of results was the motivation behind the development of the fast-guess model.

To complicate matters more, it is not always the case that the same ordering holds for both responses. For instance, both Snodgrass et al. (1967) and Carterette, Friedman, and Cosmides (1965) found RT orderings of the form $\bar{T}_{OO} < \bar{T}_{DO} < \bar{T}_{OD} < \bar{T}_{DD}$, where the stimuli have been relabeled to aid comparison. Thus, when responding "O," observers were faster when they were correct, but when responding "D," incorrect responses were faster.

Clearly, by manipulating stimulus parameters, it is possible to obtain a great diversity of results in this type of paradigm. It would therefore be a mistake to settle on a model that consistently predicts the same RT ordering in all stimulus conditions, no matter what that ordering is. Such a model might very well reflect processing structure in certain experimental conditions, but it will be inadequate as a complete model of the processing required in this experimental paradigm.

Stimulus \ Response	O	D
O	70	30
D	40	60

Stimulus \ Response	I	L
I	74	26
L	35	65

Fig. 9.2. Mean accuracies in percentages (Townsend & Snodgrass 1977).

Accuracy orderings will not be as critical in our investigations. Most experimenters arrange the stimulus conditions so that accuracy is above chance level, thus guaranteeing that correct probabilities are greater than incorrect probabilities. For instance, the accuracy data reported by Townsend and Snodgrass (1977) is illustrated in Fig. 9.2. There it can be seen that the orderings are $P_{OO}>P_{DD}>P_{DO}>P_{OD}$ and $P_{II}>P_{LL}>P_{LI}>P_{IL}$, where $P_{ij}=P(R_j\,|\,S_i)$. Notice that the accuracy orderings are the same for both target sets, with the stimuli *I* and *O* being the most recognizable.

Varied-state models

Parsimony suggests we begin our investigation of varied-state models with the simplest representatives and then later propose more elaborate models when it is necessary to do so. A very simple class of varied-state models are those that postulate only two states and allow response latencies to depend on the occupied state but not on the stimulus or the response.

We can characterize all models of this class with the following assumptions:

(A1) Upon stimulus presentation the observer is in one of two states. He or she is in state 1 with probability r and in state 2 with probability $1-r$.

(A2) If the observer is in state 1 when the stimulus is presented, mean RT will be L, but if the observer is in state 2, the mean RT will be $L+M$.

Any varied-state model postulating two internal states satisfies assumption

(A1) but may or may not satisfy (A2). Thus, an experiment falsifying the class of models we are considering here does not necessarily falsify all two-state varied-state models. It may be possible to reformulate assumption (A2) so that a more general class of two-state models results.

In spite of the mathematical simplicity of assumptions (A1) and (A2), they describe a class of models that can, conceptually, be very different depending on how one wants to define the two states. One possibility is to define a stimulus-controlled and a guessing state. In this case r is the probability that the stimulus is perfectly recognized and L is the mean response latency in these instances. With probability $1-r$ the observer extracts *no* information about the stimulus and therefore must guess between the two (in this case) alternatives. Let g equal the probability that the observer guesses stimulus A (given stimuli A and B, say) when in this guessing state and assume the act of guessing adds latency M to the total response time.

This is a simple all-or-none model (Townsend 1971a, 1978; Broadbent 1967) with no partial information states. The observer either has perfect stimulus information and is always correct or has no information and must guess randomly. The magnitude of the parameter M, which is the added latency due to guessing (i.e., response latency on guessing trials is $L+M$), determines whether the model is a fast-guess or slow-guess model. If M is negative, guesses will be, on the average, faster than stimulus-controlled responses, whereas if M is positive, guesses will tend to be slower.

Another possibility is to propose, as Falmagne (1965) has done, that the two states refer to whether the observer is prepared or unprepared for the stimulus. On the other hand, Atkinson and Juola (1974) proposed a two-state model of memory search in which the target item elicits either a very fast response (if it is either very familiar or very unfamiliar) or it initiates an extended memory search.

All of these are examples of varied-state models with two internal states. Thus, if we can falsify models satisfying assumption (A1), we will have falsified all of these models.

Our investigations of models satisfying assumptions (A1) and (A2) will be enlivened if we adopt one of these interpretations of the internal states. Since the choice is arbitrary, let us assume a stimulus-controlled and a guessing state. This choice is most consonant with the historically important notion of fast guessing.

To facilitate comparison with our earlier discussion of common empirical findings we will assume that the two stimuli to be presented are the *O* and the *D* of the Townsend and Shodgrass (1977) study. Thus we can let g be the probability the observer guesses stimulus *O* when in the guessing state. Then

$$\bar{T}_{OO}=E(\mathbf{T}\,|\,S_O,R_O)=\frac{E(\mathbf{T}\cap R_O\,|\,S_O)}{P(R_O\,|\,S_O)}$$

The numerator will contain two terms, one for the stimulus-controlled state and one for the guessing state. The stimulus-controlled mean latency is L,

the mean guessing latency is $L+M$ and the guessing state is entered with probability $1-r$. However, for this state we must also include a factor of g since the observer could be in the guessing state and not respond O. Putting these together we arrive at the second term in the numerator, $(1-r)g(L+M)$. We now have everything necessary to write $\bar{T}_{OO}$:

$$\bar{T}_{OO}=\frac{rL+(1-r)g(L+M)}{r+(1-r)g}=L+\frac{(1-r)g}{r+(1-r)g}M \tag{9.1}$$

Similarly,

$$\bar{T}_{DO}=\frac{(1-r)g(L+M)}{(1-r)g}=L+M \tag{9.2}$$

$$\bar{T}_{DD}=\frac{rL+(1-r)(1-g)(L+M)}{r+(1-r)(1-g)}=L+\frac{(1-r)(1-g)}{r+(1-r)(1-g)}M \tag{9.3}$$

and finally

$$\bar{T}_{OD}=\frac{(1-r)(1-g)(L+M)}{(1-r)(1-g)}=L+M \tag{9.4}$$

We see immediately that no matter what the values of L and M are, the model always predicts the two error latencies $\bar{T}_{DO}$ and $\bar{T}_{OD}$ to be identical. None of the studies we surveyed at the beginning of this chapter found this result. In fact, Townsend and Snodgrass (1977) found the correct latency $\bar{T}_{DD}$ to be between these two, that is, they found $\bar{T}_{DO}<\bar{T}_{DD}<\bar{T}_{OD}$.

This finding, by itself, is enough to exclude this two-state model as a complete descriptor of human choice RT performance. In fact, as noted earlier, it excludes all models satisfying assumptions (A1) and (A2). On the other hand, as was also noted, it does not necessarily rule out all varied-state models postulating two internal states. We may be able to weaken assumption (A2) in such a way that a greater diversity of latency orderings can be predicted. One possibility is to allow the latency L to depend on the ultimate decision.

The idea here is that the ultimate decision has an influence denoted by L_i ($i=O$ or D), but in the guessing state this is augmented or decremented by M. This adds one parameter to the model, since now instead of L we have L_O and L_D. The conditional latencies now become

$$\bar{T}_{OO}=\frac{rL_O+(1-r)g(L_O+M)}{r+(1-r)g}=L_O+\frac{(1-r)g}{r+(1-r)g}M$$

$$\bar{T}_{DO}=L_O+M$$

$$\bar{T}_{DD}=\frac{rL_D+(1-r)(1-g)(L_D+M)}{r+(1-r)(1-g)}=L_D+\frac{(1-r)(1-g)}{r+(1-r)(1-g)}M$$

and

$$\bar{T}_{OD}=L_D+M$$

The generality of the model has been greatly increased at the fairly inexpensive cost of one extra parameter. Depending on the values of the parameters, almost any ordering can be predicted. If M is positive (a slow-guess model), correct latencies will be faster than errors, and if it is negative (i.e., fast-guess), errors will be faster. Further, if $L_O<L_D$, then O responses will tend to be faster than D responses, whereas if $L_D<L_O$, the reverse situation occurs.

The only ordering that cannot be predicted by this class of models is the ordering $\bar{T}_{OO}<\bar{T}_{DO}<\bar{T}_{OD}<\bar{T}_{DD}$ found by Snodgrass et al. (1967) and by Carterette et al. (1965), but we will see that this type of ordering provides a dilemma for almost all choice RT models (for an exception, see the next chapter). The problem is that for response O, correct latencies are faster than errors, but for response D, error latencies are faster. In the present model, the former requires M to be positive, whereas the latter result requires M to be negative. The obvious ploy of shifting the latency differences from L to M (so that the latency parameters are L, M_O, M_D) still does not work. In this case, the model can predict $\bar{T}_{OO}<\bar{T}_{DO}$ and $\bar{T}_{OD}<\bar{T}_{DD}$, but it now predicts $\bar{T}_{OD}<\bar{T}_{DO}$ instead of the reverse ordering found in the above studies. On the other hand, such a model is as versatile as the earlier L_O, L_D, M model in the sense that it can predict as many different RT orderings. We thus have a choice of whether to place the latency differences in the stimulus-controlled or in the guessing state (or alternatively in the prepared or unprepared state if these names are preferred).

Models that postulate a latency difference in one of the two states are still two-state models since they all satisfy assumption (A1); that is, for any particular stimulus–response combination, the observable RT is still a binary mixture of response times. Thus, we see that on the basis of predicted mean RT orderings in a two-stimulus–two-response situation, we usually will not be able to rule out varied-state models that postulate only two internal states. A more sophisticated analysis is needed.

Testing the two-state hypothesis

As we shall shortly learn, the assumption that the variability in an observer's performance in a choice RT experiment is the result of mixing two distinct types of response is strong enough to allow some fairly rigorous tests to be made. These are tests of the two-state hypothesis itself, and are not meant to decide among several competing two-state models. For this reason, it will behoove us to formulate a slightly more general (i.e., less model-specific) characterization of two-state models.

Since some of the tests we will examine in this section deal with the RT density function, it makes sense to begin at this level. Now the observable RT density function $g(t)$ of a two-state model is a binary mixture of the times of the two states. Let $f_i(t)$, for $i=1$ or 2, be the RT density function for those trials in which the observer is in state i. Then

$$g(t)=pf_1(t)+(1-p)f_2(t) \tag{9.5}$$

where p is the probability of being in state 1.

At this point in our development it might not be immediately obvious how Eq. 9.5 relates to the mean RT predictions we have already derived (e.g., in Eqs. 9.1–9.4), since it does not contain a stimulus or a response subscript. In the two-state models we examined, errors occur only when the observer is in the guessing state, and so trials on which an error occurs are not assumed to be binary mixtures, as are correct response times. Thus, in the models we considered, Eq. 9.5 is assumed to hold for all correct responses. For instance, in Eq. 9.1,

$$p=\frac{r}{r+(1-r)g}$$

whereas L would be the mean of $f_1(t)$ and $L+M$ the mean of $f_2(t)$.

Tests involving mean RT

The two-state hypothesis implies several testable predictions involving mean RTs for the case when experimental manipulations induce changes in the state occupation probabilities (i.e., in p). For example, Falmagne, Cohen, and Dwivedi (1975) have shown that mean RT and the RT variance are quadratically related for all two-state models.

Proposition 9.1: If the observable RT density function is

$$g(t)=pf_1(t)+(1-p)f_2(t)$$

then under experimental manipulations of the binary mixture proportion p, the mean RT, $E(\mathbf{RT})$, and the variance of RT, $\text{var}(\mathbf{RT})$, are related by

$$\text{var}(\mathbf{RT})=-[E(\mathbf{RT})]^2+A\cdot E(\mathbf{RT})+B$$

where A and B are constants.

Proof: Let us denote the mean and variance of $f_i(t)$ by μ_i and σ_i^2. It is clear from Eq. 9.5 that

$$E(\mathbf{RT})=p\mu_1+(1-p)\mu_2 \tag{9.6}$$

and

$$\begin{aligned}\text{var}(\mathbf{RT})&=E(\mathbf{RT}^2)-[E(\mathbf{RT})]^2\\&=[pE(\mathbf{T}_1^2)+(1-p)E(\mathbf{T}_2^2)]-[p\mu_1+(1-p)\mu_2]^2\\&=\{p[E(\mathbf{T}_1^2)-\mu_1^2]+(1-p)[E(\mathbf{T}_2^2)-\mu_2^2]\}\\&\quad-[p^2\mu_1^2+2p(1-p)\mu_1\mu_2+(1-p)^2\mu_2^2]+p\mu_1^2+(1-p)\mu_2^2\\&=p\sigma_1^2+(1-p)\sigma_2^2+p\mu_1^2(1-p)-2p(1-p)\mu_1\mu_2+(1-p)\mu_2^2p\\&=p\sigma_1^2+(1-p)\sigma_2^2+p(1-p)(\mu_1^2-2\mu_1\mu_2+\mu_2^2)\\&=p\sigma_1^2+(1-p)\sigma_2^2+p(1-p)(\mu_1-\mu_2)^2\end{aligned} \tag{9.7}$$

Solving Eq. 9.6 for p and substituting this expression into Eq. 9.7 yields the following relationship between $\text{var}(\mathbf{RT})$ and $E(\mathbf{RT})$:

$$\begin{aligned}\text{var}(\mathbf{RT})&=\left[\frac{E(\mathbf{RT})-\mu_2}{\mu_1-\mu_2}\right]\sigma_1^2+\left[\frac{\mu_1-E(\mathbf{RT})}{\mu_1-\mu_2}\right]\sigma_2^2+[\mu_1-E(\mathbf{RT})][E(\mathbf{RT})-\mu_2]\\&=-[E(\mathbf{RT})]^2+\left[\mu_1+\mu_2+\frac{\sigma_1^2-\sigma_2^2}{\mu_1-\mu_2}\right]E(\mathbf{RT})+\left[\frac{\mu_1\sigma_2^2-\mu_2\sigma_1^2}{\mu_1-\mu_2}-\mu_1\mu_2\right]\end{aligned} \tag{9.8}$$

Thus the observable RT variance, $\text{var}(\mathbf{RT})$, is a quadratic function of the observable mean $E(\mathbf{RT})$. □

It should be stressed that a plot of mean RT vs. RT variance is obtained by experimental manipulations that induce systematic variations in the binary mixture proportion p. Implicit in the structure of two-state RT theories is that such manipulations do not affect the mixture densities $f_1(t)$ and $f_2(t)$. Thus the parameters μ_i and σ_i^2, $i=1,2$, are assumed constants in the above expression.

Proposition 9.1 implies that if a two-state model is correct, a plot of $\text{var}(\mathbf{RT})$ vs. $E(\mathbf{RT})$ rises to a maximum and then decreases if $A>0$ and monotonically decreases if $A\leqslant 0$. If the experimental manipulations produce large changes in p, then the sampled behavior should be diverse enough to include the inflection point. Therefore, a monotonic increasing $\text{var}(\mathbf{RT})$ vs. $E(\mathbf{RT})$ plot would be evidence against a two-state theory.

A second test of the two-state hypothesis involving mean RT is due to Yellott (1967, 1971). He showed that, given stimuli S_i and S_j, two-state models predict

$$\frac{P_{ii}\bar{T}_{ii}-P_{ji}\bar{T}_{ji}}{P_{ii}-P_{ji}}=\mu_1$$

and therefore that

$$P_{ii}\bar{T}_{ii}-P_{ji}\bar{T}_{ji}=\mu_1(P_{ii}-P_{ji}) \tag{9.9}$$

For example, from Eqs. 9.1–9.4 we see that

$$\begin{aligned}\frac{P_{OO}\bar{T}_{OO}-P_{DO}\bar{T}_{DO}}{P_{OO}-P_{DO}}&=\frac{P_{DD}\bar{T}_{DD}-P_{OD}\bar{T}_{OD}}{P_{DD}-P_{OD}}\\&=\frac{rL}{r}=L\end{aligned}$$

It will be recalled that L is the mean latency of the stimulus-controlled state.

The test of the two-state hypothesis arises when experimental manipulations produce across-condition changes in the binary mixture proportion p. On such occasions μ_1 remains invariant and so a plot of $P_{ii}\bar{T}_{ii}-P_{ji}\bar{T}_{ji}$ vs. $P_{ii}-P_{ji}$ should pass through the origin and be a linear function with slope μ_1 if the two-state hypothesis is correct. Yellott (1967, 1971) tested this predic-

tion using data sets obtained by varying payoffs and response time deadlines, factors that should affect only the binary mixture proportion p. He found that the resulting $P_{ii}\bar{T}_{ii}-P_{ji}\bar{T}_{ji}$ vs. $P_{ii}-P_{ji}$ plots were remarkably linear, just as predicted by two-state models.

Tests involving the RT density and distribution functions

In addition to the tests based on mean RT, it is possible to devise tests based on the entire RT distributions. For instance, Falmagne (1968) developed such a test, which is intended for exactly the same situations as the mean RT tests we have already examined; namely, when different state occupation probabilities occur in different experimental conditions.

Falmagne's (1968) test is in the form of a prediction called the *fixed-point property,* which he showed is true of any binary mixture of probability density functions.

Proposition 9.2 (The fixed-point property of two-state models): All binary mixtures of the same two density functions intersect at the same time point if they intersect at all.

Proof: Suppose, as in Eq. 9.5, that the density function $g(t)$ is a binary mixture of the two densities $f_1(t)$ and $f_2(t)$ (with mixture proportion p). Assume now that at some time $t=t_0$ the two densities f_1 and f_2 have the same value, say, c, so that $f_1(t_0)=f_2(t_0)=c$. This condition will be satisfied if the two densities intersect at any point $0<t_0<\infty$. Next we evaluate $g(t)$ at $t=t_0$:

$$g(t_0)=pc+(1-p)c=c.$$

Thus $g(t_0)$ is independent of p, and therefore any binary mixture of $f_1(t)$ and $f_2(t)$ must pass through the point $g(t_0)=c$. □

This property suggests that to test a two-state model we should induce the observer to change his p values over conditions. For instance, as Yellott did (1967, 1971), we might vary the speed–accuracy emphasis by means of payoffs or deadlines. The simple fast-guess model predicts that, in general, greater speed results from a higher proportion of fast guesses, that is, from an increase in p. We need only plot the RT density functions associated with each of these conditions and note their point(s) of intersection. If a two-state theory is correct, there should be one point of intersection common to all densities [i.e., the point $g(t_0)=c$]. A hypothetical situation that supports a two-state interpretation (by obeying the fixed-point property) is shown in Fig. 9.3. It should be noted, however, that tests of the fixed-point property on choice RT densities have generally failed to support two-state theories (Falmagne 1968; Noreen & Burns 1976; Lupker & Theios 1977).

More recently, Link, Ascher, and Heath (1972) generalized the fixed-point property by showing that all binary mixtures satisfy the following property.

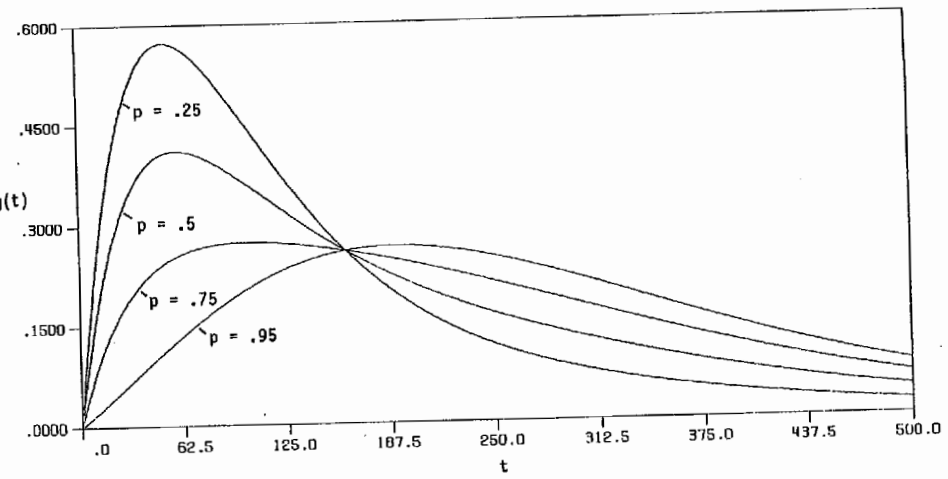


Fig. 9.3. Example of a set of RT densities satisfying the fixed-point property. The density functions are all binary mixtures of two gamma densities, one with three stages (i.e., $n=3$) and rate .01 and one with two stages and rate .02. $g(t)=pf_1(t)+(1-p)f_2(t)$, $f_1=\Gamma(n=3,v=.01)$, and $f_2=\Gamma(n=2,v=.02)$, where Γ denotes the gamma density function.

Proposition 9.3: The ratio of differences between binary mixture density or distribution functions is always constant; that is, if

$$g_i(t)=p_if_1(t)+(1-p_i)f_2(t)$$

and $G_i(t)$ is the associated distribution function, then

$$\frac{g_1(t)-g_2(t)}{g_3(t)-g_4(t)}=\frac{G_1(t)-G_2(t)}{G_3(t)-G_4(t)}=\frac{p_1-p_2}{p_3-p_4}$$

Proof:

$$\frac{g_1(t)-g_2(t)}{g_3(t)-g_4(t)}=\frac{p_1f_1(t)+(1-p_1)f_2(t)-p_2f_1(t)-(1-p_2)f_2(t)}{p_3f_1(t)+(1-p_3)f_2(t)-p_4f_1(t)-(1-p_4)f_2(t)}$$

$$=\frac{(p_1-p_2)[f_1(t)-f_2(t)]}{(p_3-p_4)[f_1(t)-f_2(t)]}=\frac{p_1-p_2}{p_3-p_4}$$

The proof for distribution functions proceeds in an analogous fashion. □

Corollary to Proposition 9.3: The fixed-point property follows from Proposition 9.3 as a special case.

Proof: Suppose $p_1=p_3$ so that $g_1(t)=g_3(t)$. Then from above,

$$[g_1(t)-g_2(t)]=\left(\frac{p_1-p_2}{p_1-p_4}\right)[g_1(t)-g_4(t)]$$

The left-hand side equals zero whenever the right-hand side equals zero. Therefore, if there exists a t_0 such that $g_1(t_0)=g_2(t_0)$, then it follows that $g_1(t_0)=g_4(t_0)$ for every value of $p_4 \neq p_1$. □

Proposition 9.3 can be used to test the two-state hypothesis because if a two-state model is correct, a plot of $G_1(t)-G_2(t)$ vs. $G_1(t)-G_3(t)$ should be linear with slope $(p_1-p_2)/(p_1-p_3)$ and zero intercept. By varying response time deadlines, Link et al. (1972) tested this condition as well as the Eq. 9.9 prediction relating mean RTs and response probabilities that was derived by Yellott (1967, 1971). They found strong support for the Eq. 9.9 prediction but even stronger evidence *against* linearity of the $G_1(t)-G_2(t)$ vs. $G_1(t)-G_3(t)$ plot. Thus, in this initial test, two-state models appear to be supported when only mean statistics are used, but when the entire RT distribution is examined they seem to fare much worse.

This finding illustrates the very important principle in stochastic modeling that testability tends to be much greater at the distributional level than it is at the mean level. Density and distribution functions carry more information than means. Thus it is not surprising that more rigorous tests of a model are possible when RT density and distribution functions are used than when only mean RT is used.

The fixed-point property can also be used as a partial test of Sternberg's finite-state models that we examined in the last chapter (Townsend 1974a). In this case, the property applies to situations where the number of elements is manipulated. Let us examine the case when two elements are to be processed, one of which is the target. Assume that with probability p the observer completes processing on the left element before the right and with probability $1-p$ completes processing on the right element before the left. Therefore, when the target is in the left-hand position it will be completed first with probability p and second with probability $1-p$. If we represent the RT density when the target is the leftmost element in a two-element display as $g_{L2}(t)$ and let $g_{R2}(t)$ be the comparable density when the target is on the right (both of which are, of course, observable), then the TP, AO, and NS assumptions of the preceding chapter (tight packing, accessibility ordering, and nested states) imply

$$g_{L2}(t)=pf_1(t)+(1-p)f_2(t)$$

where $f_i(t)$ $(i=1, 2)$ is the density associated with an item in state i ($i=1$: processed first; $i=2$: processed second). By similar reasoning,

$$g_{R2}(t)=(1-p)f_1(t)+pf_2(t)$$

We now apply the fixed-point property. If for some $t=t_0$, $f_1(t_0)=f_2(t_0)=c$, then $g_{L2}(t_0)=g_{R2}(t_0)=c$. This, in itself, tells us little. However, recall that with only one element in the display, $g_1(t)=f_1(t)$. That is, the state 1 density always applies. Therefore, $g_1(t_0)=g_{L2}(t_0)=g_{R2}(t_0)=c$. That is, the density at $n=1$ and the two serial position densities when $u=2$ are predicted to all cross at the same point. This is a prediction that is easily tested.

A possible criticism of these fixed-point arguments is that we have neglected the residual "dead" time or the base time from our analysis. This line of reasoning suggests that what we really have going on when two elements are in the display is a binary mixture of the random time to process position 1 and the time to process position 2 added to the random residual base time (e.g., stimulus encoding, response selection, and response execution). That is, we have

$$g(t)=[pf_1(t)+(1-p)f_2(t)]*b(t)$$

where $b(t)$ is the base time probability density function and $*$ denotes convolution as always (see Chapter 3). The question of interest is, of course, if the fixed-point property and the other properties we examined still hold under these conditions. A reassuring answer is easily solicited. Since convolution is distributive over addition, $g(t)$ above becomes

$$g(t)=p[f_1(t)*b(t)]+(1-p)[f_2(t)*b(t)]$$

Because $f_i(t)*b(t)$ for $i=1, 2$ is a well-defined probability density function, we see that $g(t)$ is still a binary mixture of probability densities and thus all the fixed-point properties we have discussed still hold.

We close this section with a brief look at what may be the most widely known but most misunderstood test of two-state models.

The bimodal argument against two-state models

One line of reasoning that has caused more than one two-state model to be prematurely rejected is that the observable RT density function of a two-state model should be bimodal since it is a probability mixture of two (usually) unimodal densities. Observable RT density functions are rarely found to be bimodal, and this is often taken as evidence against two-state models (e.g., Ball & Sekuler 1980).

In general, however, a bimodal mixture of two unimodal densities is the exception rather than the rule. To see this, note that if we take derivatives of both sides of Eq. 9.5, we arrive at

$$\frac{dg(t)}{dt}=p\frac{df_1(t)}{dt}+(1-p)\frac{df_2(t)}{dt}$$

Therefore, $g(t)$ will have a mode whenever

$$p\frac{df_1(t)}{dt}+(1-p)\frac{df_2(t)}{dt}=0$$

If m_1 and m_2 represent the modes of $f_1(t)$ and $f_2(t)$, respectively, then, of course (assuming continuity of the derivatives),

$$\left.\frac{df_1(t)}{dt}\right|_{t=m_1}=0 \quad \text{and} \quad \left.\frac{df_2(t)}{dt}\right|_{t=m_2}=0$$

Thus, in general,

$$\left.\frac{dg(t)}{dt}\right|_{t=m_1} = (1-p)\left.\frac{df_2(t)}{dt}\right|_{t=m_1} \neq 0$$

and

$$\left.\frac{dg(t)}{dt}\right|_{t=m_2} = p\left.\frac{df_1(t)}{dt}\right|_{t=m_2} \neq 0$$

Therefore, $g(t)$ will not typically have a mode at either m_1 or m_2.

Of course, $g(t)$ may have two modes between m_1 and m_2, but this turns out to be unlikely – in fact, impossible for certain classes of distributions. For instance, suppose $f_1(t)$ and $f_2(t)$ are unimodal and symmetric and that $f_2(t)$ has the exact shape of $f_1(t)$ but is displaced to the right or left (a common set of asumptions – e.g., Ball & Sekuler 1980). This is the situation that would result if f_1 and f_2 were both normal with the same variance but different means. This set of assumptions leads to the following result.

Proposition 9.4: If $g(t)=\frac{1}{2}f_1(t)+\frac{1}{2}f_2(t)$, where $f_1(t)$ and $f_2(t)$ are unimodal and symmetric density functions with $f_1(t)=f_2(t+a)$ for some constant a, then

(i) $g(t)$ has a mode at $t=(m_1+m_2)/2$, that is, midway between the modes of f_1 and f_2; and

(ii) it is impossible that there exist an even number of modes, so that bimodality is ruled out.

Proof: (i) We can characterize these assumptions with the following two equations: First, since f_1 and f_2 are *symmetric,* it follows that

$$f_i(x+m_i)=f_i(m_i-x), \quad i=1,2$$

and second since f_2 is a shifted version of f_1, $f_1(t)=f_2(t+a)$. Now because $m_2-m_1=a$, it follows that $f_1(x+m_1)=f_2(x+m_2)$. The two conditions together imply $f_1(m_1-x)=f_2(x+m_2)$. From this equation it follows that

$$-\frac{df_1(m_1-x)}{dx} = \frac{df_2(x+m_2)}{dx} \tag{9.10}$$

Now, $g(t)$ will have a mode whenever

$$\frac{dg(t)}{dt} = \frac{1}{2}\left[\frac{df_1(t)}{dt} + \frac{df_2(t)}{dt}\right] = 0$$

that is, if and only if

$$-\frac{df_1(t)}{dt} = \frac{df_2(t)}{dt}$$

It therefore follows from Eq. 9.10 that $g(t)$ will have a mode when $x+m_2=m_1-x$, or equivalently when $x=(m_1-m_2)/2$. Substituting this value of x back into Eq. 9.10, we conclude that $g(t)$ has a mode at $t=(m_1+m_2)/2$, that is, at the midpoint of the f_1 and f_2 modes. Also we see that at this point

$$f_1\left(\frac{m_1+m_2}{2}\right) = f_2\left(\frac{m_1+m_2}{2}\right)$$

or that the f_1 and f_2 densities intersect there.

(ii) We established above that $g(t)$ will have a mode at $t=(m_1+m_2)/2$. In addition, it will have no modes in the intervals $(0, m_1]$ and $[m_2, \infty)$ (because f_1 and f_2 are both increasing or both decreasing in these intervals). We have not ruled out the possibility of a mode in the interval $(m_1, (m_1+m_2)/2)$ but by symmetry if there is one, there must also be a mode in the interval $((m_1+m_2)/2, m_2)$. The observable density $g(t)$ will therefore have an odd number of modes all between m_1 and m_2. It therefore will never be bimodal. □

If the mixture *is* multimodal, because there probably will not be a great distance between the modes, it seems unlikely that an empirical estimate of $g(t)$ will clearly exhibit the several modes. It is more likely that $\hat{g}(t)$ will display a rather broad mode.

Multistate models

In some cases, more than two states will be needed to describe a set of data. For instance, Lupker and Theios (1975) and Falmagne and Theios (1969) argued that Falmagne (1968) failed to find support for a two-state model because three of his stimuli occurred so infrequently (with probabilities .06, .03, and .01) that a third state of "very unprepared" was created. This conclusion received some support when Falmagne and Theios (1969) found that a three-state model performed well in fitting the data.

We can expect varied-state models postulating more than two internal states to be quite flexible and perhaps difficult to falsify. In general, they will have a much greater diversity than the most general two-state models, so only the most carefully conceived experiment could hope to falsify them on the basis of mean RT orderings. To make things even worse, in a model-testing sense at least, no corresponding fixed-point property is known for three- (or more) state models. Thus in most cases, a choice between a multistate varied-state model and some other, conceptually different model of processing (e.g., a counting model) will have to be based on some criterion other than general mean RT orderings or fixed-point properties.

Conclusions

Varied-state models are based on the premise that no unitary model can adequately describe human information processing over a series of trials because the same set of processing mechanisms is not active on every trial. These models attempt to predict most of the variance in the data by identifying the number of processing structures involved. The submodels of any one of these structures are usually coarsely constructed in that they postulate few details about the nature of processing on any particular trial (i.e., in any

particular "state"). The models considered in this chapter are prototypical in this respect. The only details assumed for a given state are the RT and the probability correct when the observer is in that state. The models do not often speculate as to the nature of the processing mechanisms involved.

If one does not accept the assumption that several completely different sets of processing mechanisms determine performance in a choice RT task or if one wishes to study in some depth the details of the processing mechanisms involved, then counting and random walk models might be better theoretical vehicles than varied-state models.

Counting models

Before we begin our investigations, recall from our introductory comments that counting models assume pattern recognition is a process of accumulating evidence over time for the various stimulus alternatives, and that as soon as the evidence for one of these exceeds some preset criterion, that alternative is chosen as the response. For instance, when the stimuli are the letters O and D, presentation of either will cause evidence to be accumulated for both stimulus O and stimulus D. Counting models assume each of these accumulations occurs on some type of counter, and the recognition process is a race between a set of such counters, so that the stimulus is recognized as that alternative associated with the counter that accrues the required evidence in the shortest time.

This conceptualization works equally well for a discrete or continuous accrual of information. In the discrete case, we might naturally imagine that counter i is incremented by 1 as each chunk of evidence (e.g., a stimulus feature) is accumulated for stimulus S_i. In the continuous case, the counter might be incremented by 1 each time a certain amount of information has been accumulated. A useful analogy for this continuous-to-discrete conversion is the firing of a neuron. Excitation continually builds at the synapse, and when it reaches a certain critical level the neuron fires.

Each stimulus alternative, therefore, has a separate counter. Next let us postulate a separate criterion for each of these counters. Specifically, assume counter i emits a positive signal after incrementing K_i times. Thus, for example, the stimulus will be identified as an O if the O counter is incremented K_O times before the D counter is incremented K_D times. In addition, we shall assume the counters operate independently and in parallel. Incrementing one counter in no way affects the other. (This is one important distinction between counter models and random walk models, since in random walk models incrementing counter O simultaneously decrements counter D and vice versa.)

We might now wish to make some assumptions about the time between counts. Naturally we do not want to assume that the time between all counts is the same. If this were the case, then counts would accrue on all counters (both correct and incorrect) at the same rate and performance could never

improve above guessing. As accuracy improves, the rate at which the correct counter increments must increase. Let us assume that the times between counts on each counter are independent and all have the same exponential distribution but that the exponential rate may be different for different counters. Thus we assume some rate v_i, $i=O,D$, associated with the i counter such that the times between all counts accrued on counter i have density

$$f_i(t)=v_i \exp(-v_i t), \quad t>0$$

If the presented stimulus is O, we should expect $v_O > v_D$, but if the presented stimulus is D, we expect $v_D > v_O$. From this it is clear that we must allow the rates to depend, not only on the counter, but also on the particular stimulus presented. With two possible stimuli, this gives us four possible rates, namely v_{OO}, v_{OD}, v_{DD}, and v_{DO}, where the first subscript denotes the stimulus and the second denotes the counter. Often we will simplify this situation by assuming that all correct counts have the same rate and similarly for the incorrect counts: $v_{OO}=v_{DD}=v_c$ and $v_{DO}=v_{OD}=v_e$.

These assumptions allow us to view the recognition process as a race between two (with target set $\{O,D\}$) parallel and independent Poisson processes. If process O accumulates K_O counts before process D accumulates K_D counts, then the stimulus is recognized as an O. Although Poisson processes are well understood, because the response of the system depends on the simultaneous behavior of two such processes we cannot determine the behavior of the system by simply studying these two processes in isolation. Even so, through the superposition of these two processes, we can form a new combined process that contains all information about the behavior of the system. This technique is illustrated in Fig. 9.4.

From Proposition 3.9 we know this new process is also Poisson and has rate v_O+v_D and we can use it to recover much valuable information about the system. Consider first the event wherein the first count from the new process is generated on counter O at time t. This will happen only when the O counter increments 1 at time t *and* the D counter has not yet been incremented. The probability density of this first event is $v_O \exp(-v_O t)$ while the probability of the second event is $\exp(-v_D t)$ (i.e., the survivor function). In addition, we know that the probability density that the first count of the new process occurs at time t is $(v_O+v_D)\exp[-(v_O+v_D)t]$. Thus the probability that the first count of the new process is from the O counter given it is accrued at time t is

$$\frac{v_O \exp(-v_O t)\cdot\exp(-v_D t)}{(v_O+v_D)\exp[-(v_O+v_D)t]} = \frac{v_O \exp[-(v_O+v_D)t]}{(v_O+v_D)\exp[-(v_O+v_D)t]}$$

$$= \frac{v_O}{v_O+v_D}$$

which is independent of t.

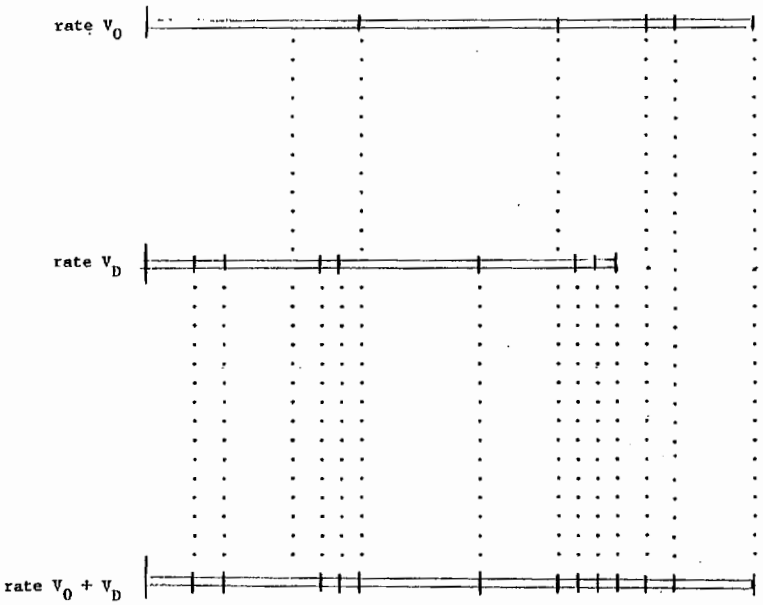


Fig. 9.4. An illustration of the superposition property of Poisson processes. If two Poisson processes with rates v_O and v_D respectively are superimposed, they form a new Poisson process with rate v_O+v_D.

A similar argument reveals that the probability that the first count arises from the D counter is $v_D/(v_O+v_D)$. By the memoryless property that characterizes all Poisson processes, we can essentially regard the process as beginning again after each count. For instance, if the first count is from counter O at time t, the approximate probability that counter O increments again in the interval $(t, t+\Delta t)$ is $v_O \Delta t$, the same as it was in every interval of that length before counter O first incremented. Similarly, the corresponding probability for counter D is $v_D \Delta t$. The Poisson memoryless property is due to the exponential interarrival time distributions and implies that no matter what point in time we might choose and no matter what the histories of the two processes are at that time, it will always be the case that the probability that the next count comes from counter O is $v_O/(v_O+v_D)$, whereas the probability it comes from counter D is $v_D/(v_O+v_D)$ until one of the counters reaches criterion.

Given the combined process, we can therefore easily calculate any desired statistic about the behavior of the system. For instance, the next result gives the unconditional response probabilities.

Proposition 9.5: In the exponential counter model with two counters, O and D, with rates v_O and v_D and with criteria K_O and K_D, the unconditional response probabilities are

$$P(R_O) = \sum_{j=0}^{K_D-1} \binom{K_O+j-1}{j} \left(\frac{v_D}{v_O+v_D}\right)^j \left(\frac{v_O}{v_O+v_D}\right)^{K_O}$$

and

$$P(R_D) = \sum_{j=0}^{K_O-1} \binom{K_D+j-1}{j} \left(\frac{v_O}{v_O+v_D}\right)^j \left(\frac{v_D}{v_O+v_D}\right)^{K_D}$$

Proof: First note that on the average, the fastest possible O response occurs when the first K_O counts on the combined process are all generated on counter O. This occurs with probability

$$\left(\frac{v_O}{v_O+v_D}\right)^{K_O}$$

The next fastest O response occurs when one D count occurs before the K_Oth O count. This D count can occur K_O different ways (i.e., before any of the counts, after the first but before the second, etc.), so the probability of this event is

$$K_O \left(\frac{v_D}{v_O+v_D}\right)\left(\frac{v_O}{v_O+v_D}\right)^{K_O}$$

Continuing in a similar manner we can compute all such probabilities because the slowest O response will occur when (K_D-1) counts have been accrued on the D counter before the K_Oth O count is recorded. The appropriate probabilities from the mean fastest to the mean slowest are given by

$$\left(\frac{v_O}{v_O+v_D}\right)^{K_O}, \quad K_O\left(\frac{v_D}{v_O+v_D}\right)\left(\frac{v_O}{v_O+v_D}\right)^{K_O},$$

$$\binom{K_O+1}{2}\left(\frac{v_D}{v_O+v_D}\right)^2\left(\frac{v_O}{v_O+v_D}\right)^{K_O}, \ldots,$$

$$\binom{K_O+K_D-2}{K_D-1}\left(\frac{v_D}{v_O+v_D}\right)^{K_D-1}\left(\frac{v_O}{v_O+v_D}\right)^{K_O}$$

Summing these all up gives the total probability of responding O. The probability of responding D is computed in a similar fashion. □

As mentioned earlier, LaBerge (1962), Audley and Pike (1965), and more recently Pike (1966, 1968, 1971, 1973) explored counter models extensively. Much of this work focused on the discrete case where no assumptions are made about the times between counts. Pike, in particular, studied the case where, after any particular unit of time, there is some probability p that one of the two counters is incremented, say counter O, and a probability $q=1-p$ that the other counter (counter D) is incremented. For our exponential case we need only set $p=v_O/(v_O+v_D)$ and $q=v_D/(v_O+v_D)$ to derive full use of Pike's results.

In a similar fashion we can compute any desired latency. For instance,

given the fastest O response (K_O counts on counter O, zero counts on counter D), the probability density that this response will take time t is

$$f(t \mid N=K_O)=\frac{[(v_O+v_D)t]^{K_O-1}}{(K_O-1)!}(v_O+v_D)\exp[-(v_O+v_D)t]$$

where N is the total number of counts accrued on the combined $O+D$ counter. This is an example of a K_O-stage gamma density with rate (v_O+v_D). The conditional mean of this density is, of course, $E(\mathbf{T} \mid N=K_O)=K_O/(v_O+v_D)$. In general, when j counts accrue on counter D (where $j<K_D$) before the K_Oth O count, the conditional probability density that response O is given at time t will be a (K_O+j)-stage gamma density with rate v_O+v_D:

$$f(t \mid N=K_O+j)=\frac{[(v_O+v_D)t]^{(K_O+j-1)}}{(K_O+j-1)!}(v_O+v_D)\exp[-(v_O+v_D)t]$$

and with conditional mean $(K_O+j)/(v_O+v_D)$.

To compute the conditional RT density function (i.e., conditioned on the counter) we must weight each possible response latency by the probability of its occurrence and then sum over all possible values of j. This leads to the following response conditioned result.

Proposition 9.6: In the exponential counter model with two counters O and D, rates v_O and v_D, and criteria K_O and K_D, the conditional processing time density functions are given by

$$f(t \mid R_O)=\frac{1}{P(R_O)}\frac{(v_O t)^{K_O-1}v_O\exp(-v_O t)}{(K_O-1)!}\sum_{j=0}^{K_D-1}\frac{(v_D t)^j\exp(-v_D t)}{j!}$$

and

$$f(t \mid R_D)=\frac{1}{P(R_D)}\frac{(v_D t)^{K_D-1}v_D\exp(-v_D t)}{(K_D-1)!}\sum_{j=0}^{K_O-1}\frac{(v_O t)^j\exp(-v_O t)}{j!}$$

and the mean conditional latencies are given by

$$\bar{T}_O=E(\mathbf{T} \mid R_O)=\frac{1}{P(R_O)}\sum_{j=0}^{K_D-1}\binom{K_O+j-1}{j}\left(\frac{v_D}{v_O+v_D}\right)^j\left(\frac{v_O}{v_O+v_D}\right)^{K_O}\left(\frac{K_O+j}{v_O+v_D}\right)$$

and

$$\bar{T}_D=E(\mathbf{T} \mid R_D)=\frac{1}{P(R_D)}\sum_{j=0}^{K_O-1}\binom{K_D+j-1}{j}\left(\frac{v_O}{v_O+v_D}\right)^j\left(\frac{v_D}{v_O+v_D}\right)^{K_D}\left(\frac{K_D+j}{v_O+v_D}\right)$$

Proof: First

$$f(t \mid R_O)=\frac{\sum_{j=0}^{K_D-1}P(R_O \cap N=K_O+j)f(t \mid N=K_O+j)}{P(R_O)} \qquad (9.11)$$

where $P(R_O)$ is given in Proposition 9.5.

Substituting into the numerator, the expression becomes

$$f(t \mid R_O)=\frac{1}{P(R_O)}\left\{\sum_{j=0}^{K_D-1}\binom{K_O+j-1}{j}\left(\frac{v_D}{v_O+v_D}\right)^j\left(\frac{v_O}{v_O+v_D}\right)^{K_O}\times\frac{[(v_O+v_D)t]^{(K_O+j-1)}}{(K_O+j-1)!}(v_O+v_D)\exp[-(v_O+v_D)t]\right\}$$

The first result can be obtained after some simplification. The derivation of $f(t \mid R_D)$ is obtained in a similar fashion.

The expressions relating the mean conditional latencies have the same form as Eq. 9.11; for example,

$$\bar{T}_O=E(\mathbf{T} \mid R_O)=\frac{\sum_{j=0}^{K_D-1}P(R_O \cap N=K_O+j)E(\mathbf{T} \mid N=K_O+j)}{P(R_O)} \quad \square$$

We are now in a position to consider the flexibility of counting models in predicting different mean RT orderings. With two stimuli and two responses, the model has six parameters in its most general formulation, four of which are rate parameters ($v_{OO}, v_{DO}, v_{DD}, v_{OD}$) and two counter criteria (K_O, K_D). With this many parameters, each free to take on any positive value, the model is very flexible and can easily predict any mean RT ordering.

In some cases, we will wish to get by with a simpler version of the model, that is, one with fewer parameters. The simplest possibility arises when $K_O=K_D=K$, $v_{OO}=v_{DD}=v_c$, and $v_{OD}=v_{DO}=v_e$. In this case both counters have the same criterion, counts on the correct counter are all accrued with rate v_c, and counts on the incorrect counter are accrued with rate v_e. Thus, there are no differential stimulus effects; for example, the behavior of the O counter when the stimulus is O is exactly the same as the behavior of the D counter when the stimulus is D. The model must predict $\bar{T}_{OO}=\bar{T}_{DD}$. A similar argument leads to the conclusion that the model predicts $\bar{T}_{DO}=\bar{T}_{OD}$. If the probability of a correct response is greater than one-half, then $v_c>v_e$ and the full set of conditional mean RT predictions becomes $\bar{T}_{OO}=\bar{T}_{DD}<\bar{T}_{DO}=\bar{T}_{OD}$. This simple counting model is incapable of predicting any of the mean RT orderings reported in the studies we reviewed at the beginning of the chapter.

An obvious generalization is to allow a different count criterion on the separate counters. Thus on counter O assume the criterion is K_O, and on counter D that it is K_D. We still assume accrual rates of v_c and v_e. This model has only one more parameter than the simple model above, but its generality is greatly increased. To see this, suppose the criterion on counter O is less stringent than the counter D criterion (i.e., $K_O<K_D$; as before, assume $v_c>v_e$); then, assuming a correct response is made, fewer counts will need to be accrued when the stimulus is O then when it is D, and so the model predicts $\bar{T}_{OO}<\bar{T}_{DD}$. By a similar argument we see that the model predicts $\bar{T}_{DO}<\bar{T}_{OD}$. On the other hand, the relation between $\bar{T}_{OO}$ and $\bar{T}_{DD}$ will

depend on the magnitude of the difference between K_O and K_D and between v_c and v_e; either mean duration can be larger.

To get a better idea as to how this model operates, suppose $v_c = .9$ and $v_e = .1$ while $K_O = 2$ and $K_D = 4$. We can now use Proposition 9.6 to compute the expected conditional mean latencies. For instance, from Eq. 9.11 we see that

$$\bar{T}_{OO} = \frac{\sum_{j=0}^{3} P(R_O \cap N = j+2 \mid S_O) E(\mathbf{T} \mid N = j+2)}{P(R_O \mid S_O)}$$

The individual terms of this expression are easily calculated. First, note that

$$E(\mathbf{T} \mid N = j+2) = \frac{j+2}{v_c + v_e} = \frac{j+2}{.9+.1} = j+2$$

while

$$P(R_O \cap N = j+2 \mid S_O) = \left(\frac{v_c}{v_c+v_e}\right)^2 \left(\frac{v_e}{v_c+v_e}\right)^j = (.81)(.1)^j$$

and

$$P(R_O \mid S_O) = \sum_{j=0}^{3} P(R_O \cap N = j+2 \mid S_O)$$

$$= .81 \sum_{j=0}^{3} (.1)^j = .999$$

Substituting these values back in and doing the same thing for the other conditional mean latencies, we find that

$$\bar{T}_{OO} = \frac{1}{.999}[(.81)2 + (.162)3 + (.0243)4 + (.003)5] = 2.22$$

$$\bar{T}_{DO} = \frac{1}{.08}[(.01)2 + (.018)3 + (.0243)4 + (.029)5] = 3.96$$

$$\bar{T}_{DD} = \frac{1}{.92}[(.656)4 + (.2624)5] = 4.27$$

and

$$\bar{T}_{OD} = \frac{1}{.0005}[(.0001)4 + (.0004)5] = 4.8$$

The resulting ordering is $\bar{T}_{OO} < \bar{T}_{DO} < \bar{T}_{DD} < \bar{T}_{OD}$, the same as reported by Townsend and Snodgrass (1977). The probabilities are also ordered in the fashion they reported: $P_{OO} = .999$, $P_{DD} = .92$, $P_{DO} = .08$, $P_{OD} = .0005$.

By manipulating the parameter values, it is also possible to achieve the ordering $\bar{T}_{OO} < \bar{T}_{DD} < \bar{T}_{DO} < \bar{T}_{OD}$. Analogously, if we had interchanged the values of K_O and K_D (so that $K_D < K_O$), then the model can predict either $\bar{T}_{DD} < \bar{T}_{OD} < \bar{T}_{OO} < \bar{T}_{DO}$ or $\bar{T}_{DD} < \bar{T}_{OO} < \bar{T}_{OD} < \bar{T}_{DO}$.

In all of these orderings, however, for a given response, correct latencies are always faster than incorrect latencies. As it turns out, this is because of

the dependence of the predicted RT orderings on the probability that an individual count is correct. For instance, assume (as would be the case in most empirical applications) that $v_c > v_e$ so that $v_c/(v_c + v_e) > \frac{1}{2}$; that is, the probability that the next count increments the correct counter is greater than one-half. It is then easy to show that correct mean latencies cannot be greater than incorrect mean latencies for any given response (i.e., the only possible orderings are $\bar{T}_{OO} \leqslant \bar{T}_{DO}$ and $\bar{T}_{DD} \leqslant \bar{T}_{OD}$). To see an intuitively related aspect of this, note that the mean fastest responses on counter i $(i = O, D)$ occur with probability

$$\left(\frac{v_c}{v_c+v_e}\right)^{K_i}$$

if the correct response is given, and with probability

$$\left(\frac{v_e}{v_c+v_e}\right)^{K_i}$$

if the incorrect response is given. But if $v_c > v_e$, then

$$\left(\frac{v_c}{v_c+v_e}\right)^{K_i} > \left(\frac{v_e}{v_c+v_e}\right)^{K_i}$$

so that the majority of the fastest responses are correct.

This class of counter models always predicts correct responses to be faster than incorrect responses, when conditioned on the response that is made. This prediction will continue to be made in the general case as long as $v_{ii} > v_{ij}$ and $v_{ii} > v_{ji}$ $(i, j = O, D)$. As noted before, this result is frequently reported in experiments employing difficult stimulus discriminations, but when discriminations are easy the faster responses tend to be errors, thereby falsifying the model. Thus, as with the two-state model treated earlier, this model cannot predict the ordering $\bar{T}_{OO} < \bar{T}_{DO} < \bar{T}_{OD} < \bar{T}_{DD}$.

At this point, we have several options available. One would be to allow the count criteria K_O and K_D to vary over trials. Pike (1973) demonstrated that if one permits sufficient variability in the criteria, then the natural mean RT order predictions of the counter model are partially or completely reversed depending upon the different parameter values. Thus, under these circumstances, counter models can predict any possible set of orderings. Although one could possibly come up with some reasonable explanation, it is not immediately clear why there should be variability in the count criteria when the discriminations are easy but not when they are difficult.

A second possibility is to assume that observers adopt different overall strategies when their task is easy. In such situations, they may become more quickly bored and begin to fast-guess on some proportion of trials. This is an example of a varied-state model (i.e., a fast-guess model) in which the processing model invoked on stimulus-controlled trials is a counting model. Thus, a prediction of this class of models is that the fixed-point property holds in cases of easy discrimination but not in cases of difficult stimulus discrimination. Of course, more complex behavior is also generated within the

class of counting models by permitting both v_e and v_c to vary without constraint across stimuli and responses.

Finer levels of analysis

The order predictions we have been considering are fairly crude in the sense that only a very low level of data analysis need be performed to determine if, say, correct latencies are faster than error latencies. Nevertheless, knowledge of the predictions of the different models at this level is very important since it allows one to quickly pare down the universe of acceptable models. But once this universe has been tentatively established, finer-grained analysis of the data will become essential. For instance, it is plausible to suppose that response time could be increased by either increasing the count criteria or decreasing the accrual rates. How are these two sets of parameters related? It is this problem to which we now turn.

Earlier we stated that the behavior of the system could not be determined by studying the individual counters in isolation. However, it turns out that we can learn at least some things about the system by *comparing* the behavior of the two individual counters. One way to do this is to establish an ordering on the appropriate finishing time distribution functions, a technique we stressed in the previous chapter. We would like to be able to show that establishing such an ordering on the isolated counters implies an ordering on the response probabilities.

Let $F_j(t)$ be the probability that counter j has accrued K_j counts by time t. More specifically, then, we would like to demonstrate that if $F_O(t) \geq F_D(t)$ for all $t>0$, then $P(\text{counter } O \text{ finishes before counter } D) \geq \frac{1}{2}$. If $\mathbf{T}_O$ is the random finishing time of counter O and $\mathbf{T}_D$ is the random finishing time of counter D, the above ordering is equivalent to saying that if $F_O(t) \geq F_D(t)$ for all $t>0$, then $P(\mathbf{T}_O < \mathbf{T}_D) \geq \frac{1}{2}$. As it happens, this ordering was established by Townsend and Ashby (1978: 227), who also showed that the reverse ordering does not hold; the response probabilities may be ordered even though the finishing time distribution functions are not. In this sense, then, an ordering of the distribution functions represents a more complete dominance of one counter over the other than does an ordering of the response probabilities.

On the other hand, even stronger dominance relations are possible. We saw some of these in the preceding chapter but we will take time to develop a few more here. Doing so will aid our investigations in at least two ways. First, it will give us several ways to demonstrate an ordering on the recognition probabilities, and second, it will allow us to determine, to a certain extent at least, the *degree* to which one counter dominates another. The full set of dominance relations that we will establish are illustrated in Fig. 9.5. The reader not interested in the mathematical details can skip down to the discussion following the proof of the counterexample to Proposition 9.8.

Implication 2 in Fig. 9.5 was established by Townsend and Ashby (1978). They showed that an ordering on the hazard functions $H_O(t)$ and $H_D(t)$

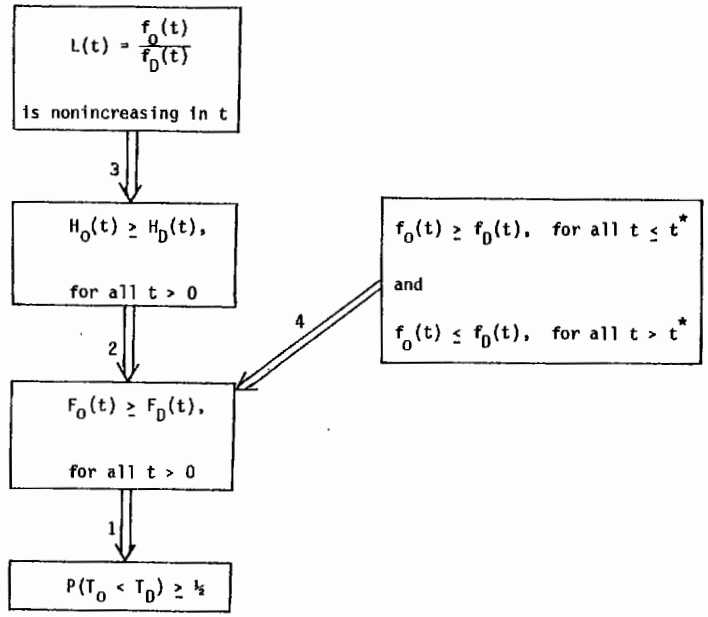


Fig. 9.5. The full set of dominance relations are established in the text. The arrows denote logical implication.

implies that the distribution functions are ordered but not vice versa. The distribution functions may be ordered even though the hazard functions are not.

Implication 3 concerns the likelihood ratio associated with the two counters O and D, which is defined as the ratio of the finishing time probability densities, namely, $L(t)=f_O(t)/f_D(t)$. When counter O is much faster than counter D, this ratio will tend to be large for small t (i.e., greater than 1) and small for large t (i.e., less than 1). Often the ratio will monotonically decrease as t increases. We now show that in such cases the associated hazard functions are ordered $H_O(t) \geq H_D(t)$ for all $t>0$.

Proposition 9.7: If $L(t)=f_O(t)/f_D(t)$ is nonincreasing as t increases, then $H_O(t) \geq H_D(t)$ for all $t>0$.

Proof: If $L(t)$ is nonincreasing, then

$$L(t)\int_t^\infty f_D(s)\,ds \geq \int_t^\infty L(s)f_D(s)\,ds \quad \text{for all } t>0$$

which is true if and only if

$$L(t)f_D(t)\int_t^\infty f_D(s)\,ds \geq f_D(t)\int_t^\infty L(s)f_D(s)\,ds \quad \text{for all } t>0$$

and hence

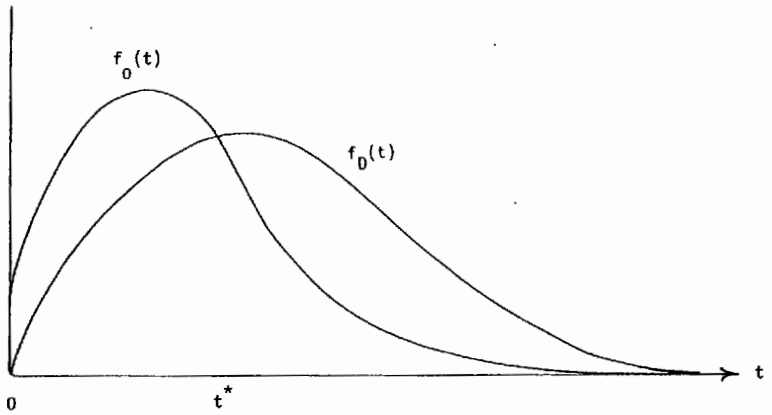


Fig. 9.6. Example of a pair of density functions for which there exists a time, $t=t^*$, such that for all $t<t^*$ one density dominates the other, but for all $t>t^*$ the dominance ordering is reversed.

$$\frac{L(t)f_D(t)}{\int_t^\infty L(s)f_D(s)\,ds} \geqslant \frac{f_D(t)}{\int_t^\infty f_D(s)\,ds} \quad \text{for all } t>0$$

Using the definition of a likelihood ratio, this inequality can be rewritten as

$$\frac{f_O(t)}{\int_t^\infty f_O(s)\,ds} \geqslant \frac{f_D(t)}{\int_t^\infty f_D(s)\,ds} \quad \text{for all } t>0$$

or finally $H_O(t) \geqslant H_D(t)$ for all $t>0$. □

Although we suspect that the converse statement is false, that is, that ordered hazard functions do not imply a nonincreasing likelihood ratio, we have not yet found a proof or a counterexample to verify our intuition.

Implication 4 in Fig. 9.5 is just one of many dominance relations we could add to our list. Most of these would be of little interest to us. For instance, many of the statistics involved are cumbersome to compute and therefore of limited empirical value. An interesting exception to this state of affairs, however, is the case where a pair of density functions cross only once. Specifically, assume as in Fig. 9.6 that $f_O(t) \geqslant f_D(t)$ for all $t \leqslant t^*$ and $f_O(t) \leqslant f_D(t)$ for all $t>t^*$. This is certainly an easy condition to verify, so if we could establish the strength of this condition, we would have an easily applied tool.

We now show that the above condition implies an ordering on the respective distribution functions.

Proposition 9.8: If there exists some time $t^*>0$ such that $f_O(t) \geqslant f_D(t)$ for all $t \leqslant t^*$ and $f_O(t) \leqslant f_D(t)$ for all $t>t^*$, then $F_O(t) \geqslant F_D(t)$ for all $t>0$.

Proof: First note that for all $t \leqslant t^*$, $F_O(t) \geqslant F_D(t)$, since $f_O(t) \geqslant f_D(t)$ over that interval. Therefore, we need only consider $t>t^*$. We wish to show that for all $t>t^*$

$$\int_0^t f_O(s)\,ds \geqslant \int_o^t f_D(s)\,ds$$

which is equivalent to showing

$$F_O(t)-F_D(t)=\int_0^{t^*} f_O(s)\,ds+\int_{t^*}^t f_O(s)\,ds-\left[\int_0^{t^*} f_D(s)\,ds+\int_{t^*}^t f_D(s)\,ds\right] \geqslant 0$$

Note that when $t=\infty$, the above is an equality ($=0$), and when $t=t^*+\epsilon$ for ϵ very small, the left-hand side is strictly greater than 0. Thus the above inequality holds if the left-hand side decreases monotonically as t increases. We can verify that the difference on the left of the inequality is monotonically decreasing by calculating its derivative. If it is always negative, the result is proven. Now

$$\frac{d}{dt}[F_O(t)-F_D(t)]=f_O(t)-f_D(t) \leqslant 0 \quad \text{for } t>t^*$$

since by assumption $f_O(t) \leqslant f_D(t)$ for all $t>t^*$. □

On the other hand, this type of ordering is not stronger than an ordering on the hazard functions. This is most easily seen via a counterexample.

Counterexample for hazard functions: If

$$\begin{aligned} f_O(t) &= 1-\tfrac{1}{2}t, && 0 \leqslant t<2 \\ &= 0 && \text{elsewhere} \end{aligned}$$

and

$$\begin{aligned} f_D(t) &= (t)^{1/2}, && 0 \leqslant t<1 \\ &= 1-(t-1)^{1/2}, && 1 \leqslant t<2 \\ &= 0 && \text{elsewhere} \end{aligned}$$

then $f_O(t)$ and $f_D(t)$ cross only once, but the hazard functions $H_O(t)$ and $H_D(t)$ are not ordered.

Proof: These two densities are shown in Fig. 9.7. As can be seen, there does exist a $t=t^*$ such that $f_O(t) \geqslant f_D(t)$ for $t<t^*$ and $f_O(t) \leqslant f_D(t)$ for $t>t^*$. As such we know that $F_O(t) \geqslant F_D(t)$ for all $t>0$. We now show that the hazard functions are not so ordered. That is, we will show that there exists at least some $t=t'$ such that $H_D(t')>H_O(t')$. Now

$$H_O(t)=\frac{1}{1-\frac{1}{2}t} \quad \text{for } 0 \leqslant t<2$$

and

$$H_D(t)=\frac{t^{1/2}}{1-\frac{2}{3}t^{3/2}} \quad \text{for } 0 \leqslant t<1$$

$$=\frac{1-(t-1)^{1/2}}{\frac{4}{3}-t+\frac{2}{3}(t-1)^{3/2}} \quad \text{for } 1 \leqslant t<2$$

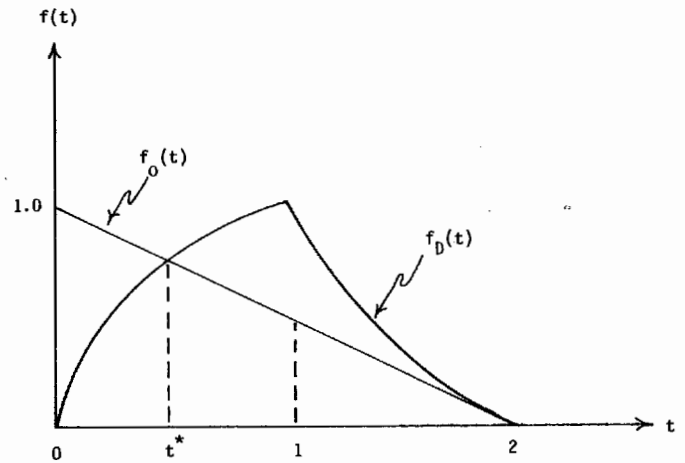


Fig. 9.7. Example of a pair of density functions, $f_O(t)$ and $f_D(t)$, for which there exists a time $t=t^*$, such that $f_O(t) \geq f_D(t)$ for all $t<t^*$ and $f_D(t) \geq f_O(t)$ for all $t>t^*$ but for which the hazard function for O is not always greater than or equal to the hazard function for D.

Substitution into these equations reveals that at $t=.9$, $H_O(t)=1.82$ and $H_D(t)=2.20$ so that an obvious crossover of the hazard functions does occur and hence they are not ordered. □

Along somewhat similar lines, Ashby (1982a) showed that if the only difference between the two densities $f_O(t)$ and $f_D(t)$, is that $f_O(t)$ includes an extra stage of processing whose duration is exponentially distributed and independent of the others, then the two densities will satisfy this property and cross only once. Further, it turns out that their point of intersection will be at the $f_D(t)$ mode. Ashby (1982a) called this the transmodal property and suggested it as a test of such models.

We need to establish one more result before continuing our investigations. So far in this section, we have concentrated on time as the variable of interest by examining how long it takes a counter to increment a fixed number of counts. Thus, $F_O(t)>F_D(t)$ for all $t>0$ means that it is always more likely that counter O has completed K_O counts than that counter D has completed K_D counts. Clearly, the probability that a counter has reached its criterion K_i $(i=O,D)$ on or before time t must equal the probability that K_i or more counts have occurred by time t. Thus, instead of focusing on the waiting time distribution associated with $F_i(t)$, one may look at the stochastic counting process formulation (see also Chapter 3).

Let $\mathbf{K}$ be the random number of counts occurring during the interval $[0,t]$. In Chapter 3 we saw that $\mathbf{K}$ has the well-known Poisson distribution,

$$P(j;t)=\frac{(v_i t)^j}{j!}\exp(-v_i t)$$

The probability that *at least* K_i counts have occurred, $P(\mathbf{K} \geq K_i; t)$, can be found by summing the $P(j;t)$,

$$P_i(K_i)=\sum_{j=K_i}^{\infty} P(j;t)$$

where we have set $P(\mathbf{K} \geq K_i; t)=P_i(K_i)$ for short. The duality between the Poisson counting process and the gamma waiting process is illustrated by the fact that

$$P_i(K_i)=\sum_{j=K_i}^{\infty}\frac{(v_i t)^j}{j!}\exp(-v_i t)=\int_0^t \frac{(v_i t')^{K_i-1} v_i \exp(-v_i t')}{(K_i-1)!}\,dt'=F_i(t) \tag{9.12}$$

This well-known identity can be derived directly by integrating by parts.

Having taken the time to build up a fairly elaborate arsenal of mathematical machinery, we are now ready to begin investigating the relationship between the different counting model parameters. The first result specifies the effects of manipulating the count criteria on the behavior of two otherwise equivalent counters.

Proposition 9.9: Suppose two Poisson counters each operate with rate $v_O=v_D=v$ but with different criteria; specifically, assume $K_O>K_D$. Then $F_O(t) \geq F_D(t)$ for all $t>0$.

Proof:

$$\begin{aligned} F_O(t)-F_D(t) &= P_O(K_O)-P_D(K_D) \\ &= \sum_{j=K_O}^{\infty}\frac{(vt)^j}{j!}e^{-vt}-\sum_{j=K_D}^{\infty}\frac{(vt)^j}{j!}e^{-vt} \\ &= \sum_{j=K_O}^{K_D-1}\frac{(vt)^j}{j!}e^{-vt} \end{aligned}$$

since $K_D>K_O$. Now $K_O \leq K_D-1$, and thus the summation is clearly greater than or equal to zero for all $t>0$, and so the result follows. □

This result is in line with our intuition. Given two counters that accrue counts at the same rate, the one with a lower criterion will always have a greater probability of finishing first (i.e., of reaching its criterion first). By the dominance relationships we therefore know that $P(\mathbf{T}_O<\mathbf{T}_D) \geq \frac{1}{2}$ and thus that the probability that the stimulus is recognized as an O is greater than the probability it is recognized as a D. The next result considers the case in which both counters have the same criterion but in which they accrue counts at different rates.

Proposition 9.10: Suppose two Poisson counters each operate with criterion $K_O=K_D=K$ but with different rates; specifically, assume $v_O>v_D$. Then $F_O(t) \geq F_D(t)$ for all $t>0$.

Proof: We will actually prove the stronger result that the likelihood ratio $L(t)$ is nonincreasing in t (see Fig. 9.5). First note that

$$L(t)=\frac{f_O(t)}{f_D(t)}=\frac{v_O\exp(-v_Ot)(v_Ot)^{k-1}/(k-1)!}{v_D\exp(-v_Dt)(v_Dt)^{k-1}/(k-1)!}$$
$$=\left(\frac{v_O}{v_D}\right)^k\exp[-(v_O-v_D)t]$$

The first term is a constant, and since $v_O>v_D$, $\exp[-(v_O-v_D)t]$ decreases monotonically and thus the likelihood ratio is nonincreasing. □

This result is also as we expected. Increasing one of the counter rates has the same qualitative effect as decreasing one of the criteria.

Finally we consider the case where the two counters differ both in rate and criterion. This case, is, in general, difficult to evaluate analytically. However, certain qualitative statements can be made immediately. We have seen that *decreasing* the criterion increases the probability that the critical count has been reached by any given time t and thus that the distribution function is increased. Similarly, *increasing* the rate also tends to increase the distribution function. It is therefore clear that simultaneously decreasing the criterion and increasing the rate will increase the distribution function and establish an ordering. The problem arises when the rate and criterion parameters are either both increased or both decreased so that their effects on the distribution function are antagonistic.

It turns out that either increasing or decreasing both the rate and the criterion does not order the distribution functions. Instead, they will always cross at some point in time. A proof of this important result will be given for completeness; however, the reader uninterested in its details may skip to the discussion that follows it.

Proposition 9.11: Suppose two Poisson counters operate with rates v_D and v_O and with criteria K_D and K_O, respectively. Suppose further that $v_D>v_O$ and $K_D>K_O$. Then there will exist finite times t_1 and t_2, both greater than zero, such that $F_O(t)>F_D(t)$ for all $t<t_1$ and $F_D(t)>F_O(t)$ for all $t>t_2$, where $F_i(t)$ is the cumulative distribution function associated with counter i.

Proof: We must prove that there always exist t_1, $t_2>0$ such that

$$F_O(t)=P_O(K_O)=\sum_{K_O}^{\infty}\frac{(v_Ot)^j}{j!}\exp(-v_Ot)\underset{\text{for } t>t_2}{\overset{\text{for } t<t_1}{\gtrless}}\sum_{K_D}^{\infty}\frac{(v_Dt)^j}{j!}\exp(-v_Dt)$$
$$=P_D(K_D)=F_D(t) \tag{9.13}$$

(a) Proof that $F_O(t)<F_D(t)$ for large t: First note that every term in the summation on the left-hand side is positive. Therefore, for any $t>0$ and any $v_O>0$, $F_O(t)$ is maximized when K_O is minimized. The smallest value K_O can take is 1. Therefore, if we can prove the contention for $K_O=1$, it must be true for all values of K_O.

Thus we wish to prove that there exists a $t_2>0$ such that for all $t>t_2$,

$$F_O(t)=\sum_{1}^{\infty}\frac{(v_Ot)^j}{j!}\exp(-v_Ot)<\sum_{K_D}^{\infty}\frac{(v_Dt)^j}{j!}\exp(-v_Dt)=F_D(t)$$

Now

$$F_O(t)=\left[\sum_{0}^{\infty}\frac{(v_Ot)^j}{j!}\exp(-v_Ot)\right]-\left[\frac{(v_Ot)^j}{j!}\exp(-v_Ot)\right]_{j=0}$$
$$=1-\exp(-v_Ot)$$

and as we expect, when the criterion is 1, the finishing time of the O counter has an exponential distribution with rate v_O. On the other hand, $F_D(t)$ is the distribution function of a K_D-stage gamma density with rate $v_D>v_O$.

A sufficient, but not necessary condition for the distribution functions to be ordered in this manner is that the tails of the density functions are ordered in the reverse direction in the same open interval. Thus $F_O(t)<F_D(t)$ for all $t>t_2$ if $f_O(t)>f_D(t)$ for all $t>t_2$. We will verify that this latter inequality holds under the present circumstances. Specifically, we wish to verify that there will always exist a $t_2>0$ such that for all $t>t_2$,

$$f_O(t)=v_O\exp(-v_Ot)>\frac{(v_Dt)^{K_D-1}}{(K_D-1)!}v_D\exp(-v_Dt)=f_D(t)$$

which is equivalent to verifying that

$$\frac{(K_D-1)!v_O}{v_D^{K_D}}>\frac{t^{K_D-1}}{\exp[(v_D-v_O)t]}\quad\text{for all } t>t_2$$

By hypotheses, $K_D>1$ whereas v_D and v_O are greater than zero; therefore, the left-hand side is a positive constant. On the other hand,

$$\lim_{t\to\infty}\frac{t^{K_D-1}}{\exp[(v_D-v_O)t]}=0$$

and so a t_2 with the desired properties must exist.

(b) Proof that $F_O(t)>F_D(t)$ for small t: Rewriting Eq. 9.13 slightly reduces our task to one of finding a $t_1>0$ such that for all $t<t_1$,

$$1-F_O(t)=\sum_{0}^{K_O-1}\frac{(v_Ot)^j}{j!}\exp(-v_Ot)<\sum_{0}^{K_D-1}\frac{(v_Dt)^j}{j!}\exp(-v_Dt)=1-F_D(t)$$

Notice that each term in the summation on the right-hand side is positive for all values of $t>0$. Therefore, if the inequality holds for a particular value of K_D, it must necessarily hold for all larger values of K_D. Because $K_D>K_O$, the smallest value it can take is $K_D=K_O+1$. Therefore we will consider this case.

Note first that $F_O(t)=F_D(t)=0$ for $t=0$. Therefore a sufficient condition for the distribution functions to be ordered in the way we require is that the densities are ordered in the same fashion. Thus, we must find a $t_1>0$, such that for all $0<t\leqslant t_1$,

$$f_O(t)=\frac{(v_O t)^{K_O-1}}{(K_O-1)!}v_O\exp(-v_O t)>\frac{(v_D t)^{K_D-1}}{(K_D-1)!}v_D\exp(-v_D t)=f_D(t)$$

Because we are assuming that $K_D=K_O+1$ this inequality becomes

$$\frac{(v_O t)^{K_O-1}}{(K_O-1)!}v_O\exp(-v_O t)>\frac{(v_D t)^{K_O}}{K_O!}v_D\exp(-v_D t)$$

A little rearranging shows that this inequality holds whenever

$$t^{K_O-1}\left[v_O^{K_O}\exp(-v_O t)-\frac{v_D^{K_O+1}}{K_O}t\exp(-v_D t)\right]>0$$

This inequality holds if we can find a t_1 such that for all $0<t\leqslant t_1$,

$$v_O^{K_O}\exp(-v_O t)>\frac{v_D^{K_O+1}}{K_O}t\exp(-v_D t)$$

or equivalently whenever

$$\frac{K_O v_O^{K_O}}{v_D^{K_O+1}}\exp(v_D-v_O)t>t$$

Now $\exp(v_D-v_O)t\geqslant 1$ because $v_D>v_O$ and so this inequality holds for all

$$t<\frac{K_O v_O^{K_O}}{v_D^{K_O+1}}\quad\square$$

To get a better intuitive feeling for the behavior of a set of counters with antagonistic criteria and accrual rates, let us consider a numerical example. Suppose that $v_O=1$ and $v_D=2$ but that $K_O=3$ and $K_D=5$. Thus the O counter accrues counts at a slower rate, but it also has a lower criterion, so that it needs fewer total counts to trigger a response. The behavior of the distribution functions associated with these two counters is illustrated in Table 9.1 over the range $t=.1$ to $t=5$. As can be seen, there is a clear crossover of the distribution functions somewhere in the interval $1.0<t<2.0$ and, as they should be, the tails are ordered as predicted by Proposition 9.11. For $t\leqslant 1.0$, counter O dominates and $F_O(t)>F_D(t)$, while for $t\geqslant 2.0$ the dominance reverses and $F_D(t)>F_O(t)$.

Thus, the O counter starts faster but finishes slower than the D counter. In other words, for very short times it is more likely that counter O has accrued 3 counts than that counter D has accrued 5 counts, even though counter D accrues counts at an overall faster rate. However, at long times the situation is reversed and counter D is more likely to be completed (i.e., the advantage of a greater rate begins to pay off). So for small t, the criterion is more important than the accrual rate in the sense that the distribution function is larger for the smaller criterion. For large t, the rate dominates, in that the distribution function associated with the larger rate is greater than the one with a smaller rate. Proposition 9.11 assures us that these conclusions depend only on the fact that $v_D>v_O$ and $K_D>K_O$ and not on the specific parameter values of the example.

Table 9.1. *Sample distribution functions of the exponential counter model*

	$v_D=2$ $K_D=5$	$v_O=1$ $K_O=3$
	$F_D(t)=\sum_{j=5}^{\infty}\frac{(2t)^j}{j!}e^{-2t}$	vs. $\sum_{j=3}^{\infty}\frac{t^j}{j!}e^{-t}=F_O(t)$
t	$F_D(t)$	$F_O(t)$
.1	0	.0002
.2	.0001	.0011
.3	.0004	.0036
.4	.0014	.0079
.5	.0037	.0144
.6	.0077	.0231
1.0	.0527	.0803
2.0	.3712	.3233
3.0	.7149	.5768
4.0	.9004	.7619
5.0	.9707	.8753

Conclusions

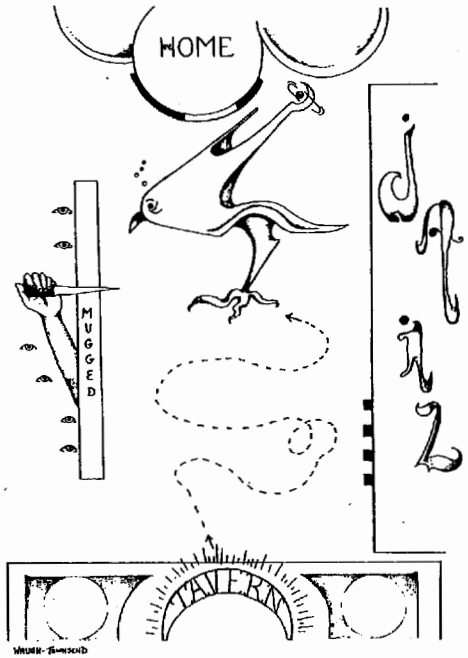
In this chapter, we first analyzed varied-state models, with an emphasis on two-state models. We learned that a model positing that in state 1 the mean latency is L_i and in state 2 it is L_i+M was quite flexible but could not predict the latency order $\bar{T}_{OO}<\bar{T}_{DO}<\bar{T}_{OD}<\bar{T}_{DD}$. The letters O and D symbolize two different stimuli ($\bar{T}_{ij}=$ mean latency conditioned on stimulus i and response j). Next a number of other tests of two-state models were discussed.

Turning to counting models, even some quite general but still falsifiable classes of this type were shown to be incapable of predicting the above ordering on $\bar{T}_{ij}$.

Finally, some fine-grained aspects of the counting distributions were examined, particularly with regard to determining distribution orderings (which in turn imply that the process associated with the larger distribution function will finish first more than one-half the time). More work needs to be done relating this level of analysis to the overall performance of two counters working in unison.

In the next chapter, an alternative but competing class of models will be considered that are also based on the notion of counting processes. The ran-

A dependence between counters is enough to radically alter the intuitive conception of the nature of processing. Thus, random walk models cannot be considered a special case of the counting models we discussed in this chapter. For this reason, they will be examined in a chapter of their own.



This chapter is devoted to a discussion of *random walk models* with two absorbing barriers. Here our slightly inebriated protagonist begins at the doors of the tavern and proceeds randomly on his way home. Two absorbing barriers are jail and "getting mugged." Home might be viewed as a third "barrier," but, as it is moved infinitely far away relative to the closeness of the other two barriers (and the "drift" of the hero), this evolves into a two-barrier situation. See text for a more sober treatment.

10 *Random walk models of reaction time and accuracy*

In this chapter we continue our investigation of choice RT models that possess enough structure to predict both accuracy and latency results. The random walk models we examine in the following pages are very much like the counter models of Chapter 9, in the sense that they both assume that stimulus information is accumulated gradually over time and that a response is made as soon as the evidence in favor of one alternative exceeds some preset criterion level.

As we shall see, however, the differences between these classes of models are great enough to warrant independent treatments. Even so, to facilitate comparison between random walk models and counting models, as well as varied-state models, the focus of this chapter, like the last, will be on the mean RT orderings that random walk models predict in two alternative forced choice tasks.

In this type of experimental paradigm an observer is presented with one of two stimuli. His or her task is to identify the presented stimulus as quickly as possible. Both counting models and random walk models assume that on

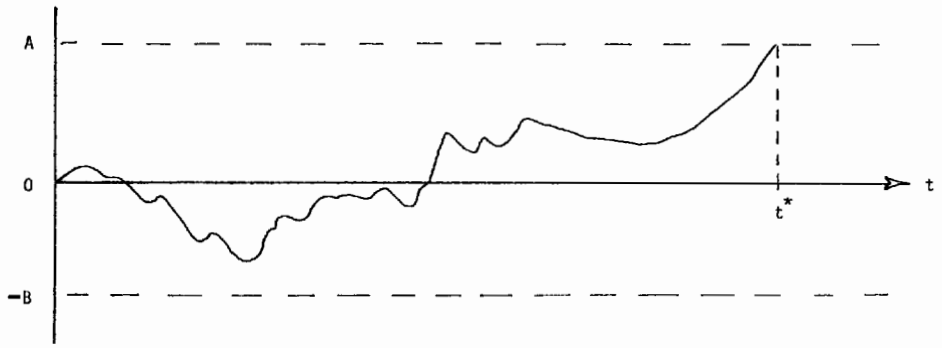


Fig. 10.1. Schematic representing a hypothetical trial as postulated by a random walk model of choice RT. The walk begins at the point zero and drifts randomly up and down over time until the barrier at A or $-B$ is crossed. In this case, response R_A is given since the barrier at A is crossed first and the discrimination time is t^*.

presentation of the stimulus the observer begins accumulating evidence in favor of each of the two stimuli. In counting models, evidence in favor of each stimulus is accumulated independently. In random walk models, however, there is a complete dependence between accumulators. This is because, in this class of models, evidence for one stimulus is simultaneously regarded as evidence *against* the other. When one of the two alternatives accumulates a certain preset amount of evidence in its favor, *relative* to the other alternative, then it is selected as the response.

Analytically, this set of assumptions describes a random walk in time with two thresholds or barriers, one for each of the response alternatives. Just prior to stimulus presentation the random walk is at the origin. When the stimulus is presented the walk moves back and forth between the barriers as evidence is collected for the two response alternatives. Eventually the walk crosses one of the two barriers and a response is given.

This situation is illustrated in Fig. 10.1. There the walk starts at the origin and the barrier associated with response R_A is at point A whereas the barrier associated with response R_B is at point $-B$. A difference in the values of A and B represents a bias toward one response alternative or the other. Discrimination time in the example of Fig. 10.1 is equal to t^*, the time when one of the two barriers is first crossed.

Response probability is determined by two factors: the values of A and B and the drift tendencies of the walk. If A equals B and the walk has a greater tendency to drift to one of the two barriers, say toward the barrier associated with response R_A, then the probability that the response is R_A will be greater than the probability that the response is R_B. Because of this, in a two-choice discrimination task where accuracy is above chance level, we would expect

the drift rates to the two barriers to depend upon the identity of the presented stimulus. When stimulus S_A is presented, the drift toward the barrier at A must be greater than the drift toward the barrier at $-B$ for accuracy to be above chance. When stimulus S_B is presented the opposite must occur: The drift toward the barrier at $-B$ must be greater.

We shall assume that the observer samples the presented stimulus (and/or memory of the stimulus) at discrete time points, Δt time units apart. Thus, movement in our random walk can occur only at time points that are multiples of Δt; that is, the time axis is assumed to be discrete. This makes calculation of discrimination time an easy matter: We simply multiply the number of steps to absorbtion by Δt. For illustrative convenience, we shall assume throughout this chapter that Δt equals 1 msec.

In contrast to the time axis, the step sizes are assumed to lie on a continuous scale, making the amount of drift occurring at any one time point a continuous but random variable. We can describe this randomness by identifying a step size probability density function. If this density function has a positive mean, the walk will typically tend to drift toward barrier A, whereas a negative mean will tend to push it toward the barrier at $-B$.

Random walk models of choice reaction time have become increasingly popular during the past 20 years or so. A large part of this popularity is undoubtedly due to the recent concern over the speed–accuracy trade-off that occurs in choice RT tasks (see Pachella 1974). The models of this and the preceding chapter are some of the very few making explicit predictions about this trade-off.

In 1960 Stone proposed a random walk model of choice reaction time based on the sequential probability ratio test (SPRT) of Wald (1947). This model has since been elaborated by Edwards (1965) and Laming (1968). It is based on the assumption that the observer makes optimal use of the information available, in the sense that to maintain the same accuracy level, any other processing strategy will lead to a greater mean RT.

More recently, Link and Heath (1975) developed a class of random walk models quite different in formulation from the SPRT model. The most general of these is difficult to test because of its many parameters. In an effort to construct a simpler model, Link and colleagues (Link & Heath 1975; Link 1975) studied one model from this family that postulates that, given barriers at A and $-B$, the drift toward A when the stimulus is S_A equals the drift toward $-B$ when the stimulus is S_B. Link and Heath call this an assumption of unbiasedness. The resulting model is tractable and it contains one more parameter than the SPRT model.

The very different foundations upon which these models were formulated has naturally caused some confusion as to their relation. A well-known prediction that has come to characterize the SPRT random walk model is that, when conditioned on the response made, correct and incorrect mean latencies must be equivalent. We saw in the preceding chapter that empirical correct and error latencies are not usually found to be equal. On the other hand,

Link and Heath (1975) showed that their random walk model can predict any ordering between correct and incorrect latencies (for a given response) and thus contains the SPRT prediction *on this statistic* as a special case.

Recently, however, Swensson and Green (1977) showed that the SPRT model is not a special case of the Link and Heath unbiasedness model. Instead, the two models intersect, but neither contains the other.

Before we get too far ahead of ourselves let us examine these models more carefully. First we note the preliminary assumptions made by both models that transform the recognition process into a random walk over time.

To begin with, both models assume that for each interval of length Δt, a psychological value is registered along some continuum. These values determine a distribution along the psychological continuum that comes to characterize the stimulus. Sampling from stimulus S_i for a time interval of length $k\Delta t$ will therefore generate a set of k psychological values. These values are assumed to be drawn from some distribution characterizing the stimulus. The recognition process now becomes equivalent to determining from which distribution a set of psychological values has arisen. Familiarity with the stimulus implies familiarity with its psychological distribution and thus, through experience with a stimulus, the observer may discover its psychological distribution. It is important to realize, however, that the true effect of the stimulus is not under the observer's control. True, this effect may depend on many factors (attention, bias, etc.), but it is presumably these factors that determine the psychological distribution associated with the stimulus.

The SPRT random walk model

The SPRT model assumes that for each observed psychological value the observer calculates the likelihood (by computing the likelihood ratio) that the value was sampled from the distribution associated with, say, stimulus S_A. This likelihood then determines the movement of the walk.

To put things a little more explicitly, let us denote the psychological continuum by y and the probability density function of the psychological values generated by stimulus S_i by $f_i(y)$. Then the SPRT model assumes that, given the registration of some particular psychological value, say, $\mathbf{Y}=\mathbf{Y}_1$, the observer calculates the statistic $\mathbf{X}_1=\ln[f_A(\mathbf{Y}_1)/f_B(\mathbf{Y}_1)]$. If the result lies above A then response R_A is given, if it lies below $-B$, response R_B is made and otherwise another sample is taken. The log likelihood of this second sample is computed and then added to $\ln[f_A(\mathbf{Y}_1)/f_B(\mathbf{Y}_1)]$ and the whole decision process is repeated.

When the likelihood ratio is greater than 1, evidence favors R_A, and when it is less than 1, R_B is more likely to be correct. Thus, the logarithmic transformation moves the walk up when the evidence supports R_A and moves it down when R_B is supported. The response criteria form absorbing barriers at the points A and $-B$. The word *absorbing* refers to the fact that once either of the barriers is crossed, the walk terminates.

This strategy of accumulating log likelihood ratios means that the position of the walk at time $k\Delta t$ (i.e., after k psychological values have been observed and assuming a barrier has not been crossed) is given by

$$\mathbf{W}_k=\sum_{j=1}^{k}\ln\left[\frac{f_A(\mathbf{Y}_j)}{f_B(\mathbf{Y}_j)}\right]=\sum_{j=1}^{k}\mathbf{X}_j$$

where $\mathbf{Y}_j$ is the random value of the jth psychological sample. Thus, the size of the jth step of the random walk is

$$\mathbf{X}_j=\ln\left[\frac{f_A(\mathbf{Y}_j)}{f_B(\mathbf{Y}_j)}\right]$$

The behavior of this random walk has been extensively studied, since it is equivalent to the well-known sequential probability ratio test (SPRT) of Wald (1947). Since $\mathbf{Y}_j$ is a random variable, so is $\mathbf{X}_j$, and since on any one trial the $\mathbf{Y}_j$ are assumed to be sampled independently from the same distribution [either $f_A(y)$ or $f_B(y)$], the $\mathbf{X}_j$ must also all be independent and identically distributed. Let us call the density function of the random step size when stimulus S_i ($i=a$ or b) is presented, $g_i(x)$. As it turns out, the two step size density functions $g_A(x)$ and $g_B(x)$, as well as the distance to the barriers A and B, completely determine the behavior of the random walk.

Relative judgment theory (RJT)

The random walk model originating from Link and Heath's (1975) relative judgment theory (RJT) involves no transformation of the psychological axis. Instead, the theory assumes that the observer solves the discrimination problem by first constructing a mental (but random) referent on the same psychological continuum. The referent acts as a standard against which the sampled stimulus values can be compared. In the most general RJT model, the only constraint placed on the referent $\mathbf{Y}_r$ is that its mean lies between the means of the two stimulus distributions. This implies that sampled values that are greater than the referent tend to favor one response, whereas values less than the referent favor the other.

More specifically, suppose a psychological value $\mathbf{Y}_j$ is sampled from the to-be-discriminated stimulus at time $j\Delta t$; then the model assumes that the observer draws a sample $\mathbf{Y}_{r_j}$ from the referent distribution and then calculates the difference $\mathbf{X}_j=\mathbf{Y}_j-\mathbf{Y}_{r_j}$. This difference determines the size of the jth step of the random walk. The position of the walk at time $k\Delta t$ is thus

$$\mathbf{W}_k=\sum_{j=1}^{k}(\mathbf{Y}_j-\mathbf{Y}_{r_j})=\sum_{j=1}^{k}\mathbf{X}_j$$

As in the SPRT model, on any one trial, the $\mathbf{Y}_j$ are sampled independently from the same distribution [either $f_A(y)$ or $f_B(y)$]. In addition, the $\mathbf{Y}_{r_j}$ are random values of the referent, sampled independently from the distribution $f_r(y)$. These facts guarantee that the random step sizes $\mathbf{X}_j$ are all independent

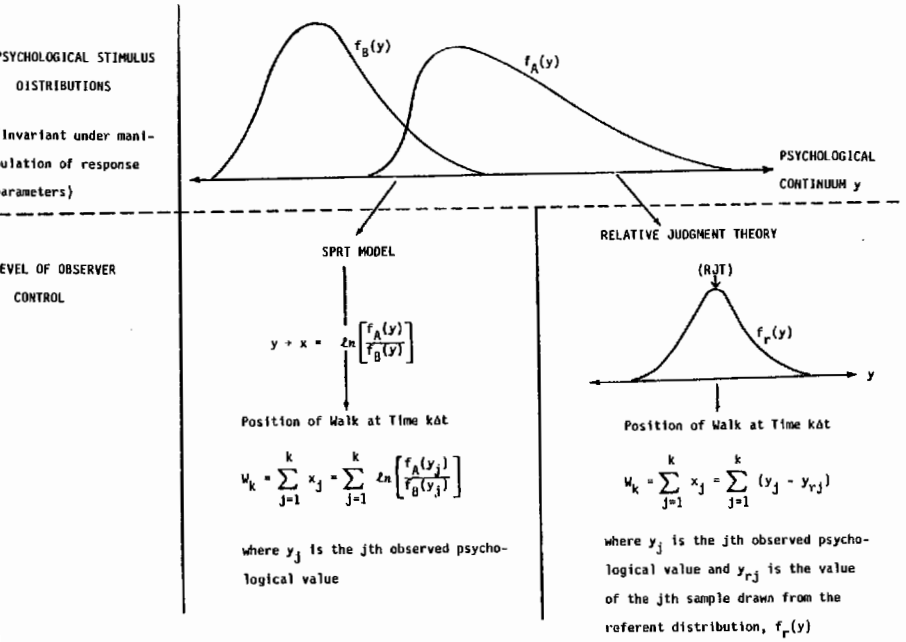


Fig. 10.2. A schematic which details the internal events that are postulated by the SPRT model and by relative judgment theory (RJT) to occur during discrimination.

and have the same sampling distribution $g_i(x)$, where the subscript i denotes the fact that for a given referent the drift of the walk depends only on which stimulus was presented. Thus, as in the SPRT model, the behavior of the RJT walk is completely determined by the step size densities $g_A(x)$ and $g_B(x)$ and by the barrier parameters A and B. The difference between the two models is in the relationship they predict to exist between $g_A(x)$ and $g_B(x)$. In the case of the SPRT model, we have not yet made the nature of this relationship clear; however, one might surmise that the two densities are related in some fashion, since both are determined by the same transformation, $\mathbf{Y} \to \ln[f_A(\mathbf{Y})/f_B(\mathbf{Y})]$. We shall investigate the nature of this relationship later in the chapter.

In the unbiasedness version of the RJT random walk model, it is assumed that the drift toward the barrier at A when stimulus S_A is presented is the same as the drift toward the barrier at $-B$ when the stimulus is S_B. This assumption fixes the relationship between the two step size density functions as $g_A(x) = g_B(-x)$. Thus the two densities are mirror images of each other.

It is important to note that this simplifying assumption of unbiasedness is aimed at a much different level of the observer's processing system than is the assumption of the SPRT model (i.e., that the observer performs the transformation $\mathbf{Y} \to \ln[f_A(\mathbf{Y})/f_B(\mathbf{Y})]$). In the RJT model (see Fig. 10.2) the observer can influence the step size densities $g_A(x)$ and $g_B(x)$ by careful selection of the internal referent, which is completely under his or her control. But since the step size is determined by subtracting a value of the referent from the observed stimulus value, the step size distributions depend also on the original stimulus distributions themselves, which are not under the observer's control. As a result, any simplifying assumption made by RJT about the form of the step size distributions is an assumption, at least to some extent, about the *initial* effect of the stimuli upon the observer's processing system. It is thus an assumption about a sensory event rather than a decision event. Since we expect these types of sensory processes to be affected only by changes in stimulus presentation and not by any changes in decisional influences or observer bias, we therefore expect any simplifying assumption such as "unbiasedness" [i.e., $g_A(x) = g_B(-x)$] to hold at least for all experimental conditions in which the stimulus event is static.

The SPRT assumption is, on the other hand, an assumption about the decision process. The observer is assumed to have learned the optimality of the transformation $\mathbf{Y} \to \ln[f_A(\mathbf{Y})/f_B(\mathbf{Y})]$ and chosen to employ it under most circumstances. No assumptions are made about the original stimulus distributions $f_A(y)$ and $f_B(y)$. As instructions to the observer change, however, his definition of optimality may also change, and with it the transformation he chooses to make. For instance, under other instructions the observer may choose to select a referent value and subtract it from the observed sensory value, as in the RJT models.

Derivation of response probabilities and mean RT statistics

Since the behavior of the SPRT and the RJT random walks are both completely determined by the step size probability density functions and the barrier parameters A and B, there is no need to derive response probabilities and mean RT predictions for each model separately. Instead, we can derive predictions for a general model assuming arbitrary step size densities and then afterward see how these predictions are constrained by the separate assumptions of the two models. An advantage of this strategy is that if some new random walk model is proposed in the future, its predictions can be derived from the predictions of the general model.

Our derivations will follow the development of Cox and Miller (1965), which depends upon certain important properties of the moment-generating function (mgf). Thus before proceeding further we will take time to review some of these properties (see also Chapter 3).

First recall that the mgf of the step size distribution when the presented stimulus is S_A is given by

$$M_A(\theta) = \int_{-\infty}^{\infty} e^{-\theta x} g_A(x)\, dx$$

Note that when $\theta = 0$, $M_A(\theta) = 1$, no matter what the form of $g_A(x)$. Thus, for all mgfs, $M_A(0) = 1$. Equally clear is the fact that $M_A(\theta) > 0$ for every

possible value of θ, since both $e^{-\theta x}$ and $g_A(x)$ are always nonnegative. If all response probabilities are nonzero, then the step size probability distributions must have measurable probability density on both sides of the origin. A technical outcome of this is that as θ approaches either $+\infty$ or $-\infty$, $M_A(\theta)$ will approach $+\infty$. It can also be shown (Cox & Miller 1965) that this approach to ∞ is monotonic in each direction. Finally, it is easy to see that $d^2M_A(\theta)/d\theta^2 > 0$; that is, $M_A(\theta)$ is positively accelerated in θ. Thus our picture of the mgf is fairly complete. It is convex (i.e., it is a U-shaped function opening upward, although it is not necessarily symmetric), it lies in the upper half-plane, and it crosses the vertical axis at the point $M_A(\theta)=1$ (see Fig. 10.3).

Next we recall the famous property from which mgfs derive their name (see Proposition 3.5). The mean of the step size distribution is equal to -1 times the slope of the tangent at the point where $M_A(0)=1$. This is because

$$\begin{aligned}\left.\frac{dM_A(\theta)}{d\theta}\right|_{\theta=0} &= \left.\frac{d}{d\theta}\int_{-\infty}^{\infty} e^{-\theta x} g_A(x)\,dx\right|_{\theta=0} \\ &= \left.\int_{-\infty}^{\infty} \frac{d}{d\theta} e^{-\theta x} g_A(x)\,dx\right|_{\theta=0} \\ &= \left.\int_{-\infty}^{\infty} -xe^{-\theta x} g_A(x)\,dx\right|_{\theta=0} \\ &= \int_{-\infty}^{\infty} -x g_A(x)\,dx = -\mu_A\end{aligned}$$

This property tells us that if the minimum of the function $M_A(\theta)$ occurs to the right of the origin (as in Fig. 10.3), then the step size mean will be positive. If the minimum occurs to the left of the origin, the mean is negative; and if the minimum occurs at $\theta=0$, the mean will be zero. Because the expected drifts are assumed to be nonzero in most random walk models of choice RT, the functions $M_A(\theta)$ and $M_B(\theta)$ will have their minimum at some point other than the point $\theta=0$. This guarantees that the equation $M_i(\theta)=1$ will have two solutions. Call these roots θ_1 and θ_2. We already know that $\theta_1=0$ is one solution. Assume the point $\theta_2=\theta_i$ (where $i=A$ or B) is the second root, so that $M_i(\theta_1=0)=M_i(\theta_2=\theta_i)=1$. The two roots of the equation $M_i(\theta)=1$ turn out to be very important to the theory of random walks. Almost as important are the derivatives evaluated at these two values of θ.

We know from above that

$$\left.\frac{dM_i(\theta)}{d\theta}\right|_{\theta=\theta_1=0} = -\mu_i$$

If the mgf $M_i(\theta)$ is a symmetric function, then

$$\left.\frac{dM_i(\theta)}{d\theta}\right|_{\theta=\theta_2} = \mu_i$$

However, if it is not symmetric, the derivative evaluated at $\theta_2=\theta_i$ will not equal μ_i. Following Link and Heath (1975), let us call the derivative at $\theta=\theta_2$

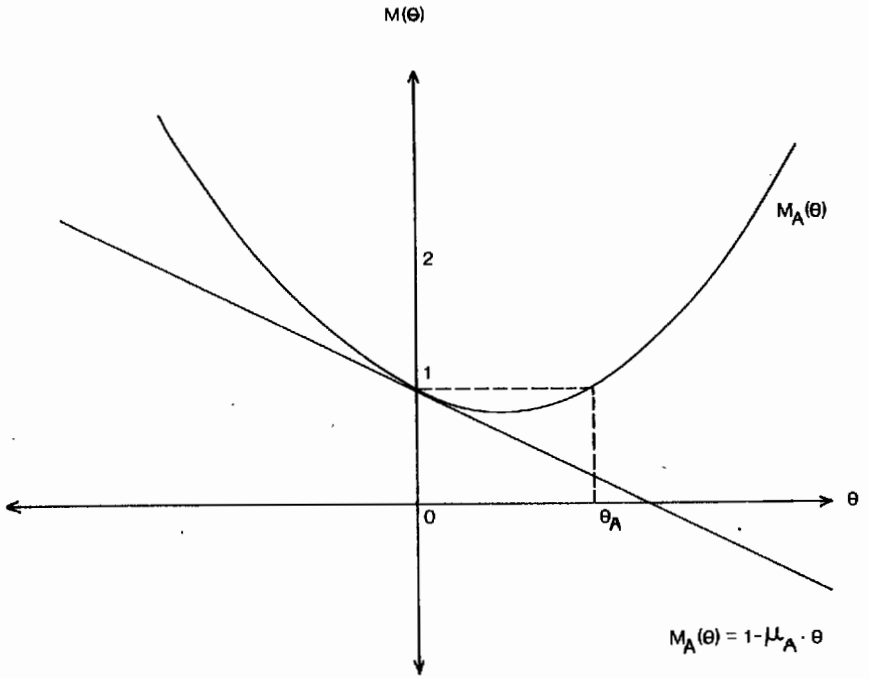


Fig. 10.3. An example of a moment generating function (mgf) of a density function with a positive mean.

$$\left.\frac{dM_i(\theta)}{d\theta}\right|_{\theta=\theta_2} = c_i\mu_i$$

If $c_i=1$, then $M_i(\theta)$ must be symmetric. If $c_i<1$, it will be skewed to the right; and if $c_i>1$, it will be skewed left. Thus the parameter c_i is an indicant of the direction of asymmetry in $M_i(\theta)$.

We are now ready to derive response probabilities and mean RT predictions of a general random walk model. The basis of the approach we will follow is Wald's (1947) fundamental identity of sequential analysis. The form of this identity most useful to us is, for the case when stimulus S_i is presented,

$$E[\exp[-\theta_j(z_i)\mathbf{W}_n]z_i^n]=1, \quad j=1,2 \tag{10.1}$$

where $\mathbf{W}_n$ is the position of the walk at the moment of absorption. Thus n is the number of steps until absorption occurs when the stimulus is S_i. The term z_i equals the reciprocal of the step size mgf, that is, $z_i=[M_i(\theta)]^{-1}$, and the $\theta_j(z_i)$ are, as before, the two roots of the equation $M_i(\theta)=1$. Thus $\theta_1(z_i)=0$ and $\theta_2(z_i)=\theta_i$, where $i=A$ or B. The dependence on z_i emphasizes the fact that the roots θ_j depend on the mgf $M_i(\theta)$. Wald's identity, as it is presented in Eq. 10.1, might not offer any immediate insight to the nonspecialist, but it turns out to be extremely useful nonetheless. With it, several critical deriva-

tions are rendered much more straightforward than would otherwise be possible. The first result lists the predictions of interest to us.

Proposition 10.1: In the most general random walk model, the error probabilities are given by

$$P(R_B \mid S_A) = \frac{1-\exp(-\theta_A A)}{\exp(\theta_A B)-\exp(-\theta_A A)}$$

and

$$P(R_A \mid S_B) = \frac{\exp(\theta_B B)-1}{\exp(\theta_B B)-\exp(-\theta_B A)}$$

whereas the predicted conditional mean RTs are approximately

$$\bar{T}_{iA} = \frac{1}{c_i \mu_i}\left[\frac{(Ac_i+B)\exp(\theta_i B)+(A+Bc_i)\exp(-\theta_i A)}{\exp(\theta_i B)-\exp(-\theta_i A)} - \frac{Bc_i+B\exp(\theta_i B)}{\exp(\theta_i B)-1}\right]$$

and

$$\bar{T}_{iB} = \frac{1}{c_i \mu_i}\left[\frac{(Ac_i+B)\exp(\theta_i B)+(A+Bc_i)\exp(-\theta_i A)}{\exp(\theta_i B)-\exp(-\theta_i A)} + \frac{Ac_i+A\exp(-\theta_i A)}{\exp(-\theta_i A)-1}\right]$$

where $i=A$ or B and, as in the preceding chapter, $\bar{T}_{ij}$ is the mean RT of response R_j when the stimulus is S_i.

Proof: We begin by expanding Eq. 10.1 so that it incorporates a conditional dependence on the response that is made,

$$P(R_A \mid S_i)E\{\exp[-\theta_j(z_i)\mathbf{W}_n]z_i^n \mid \mathbf{W}_n \geqslant A\} + P(R_B \mid S_i)E\{\exp[-\theta_j(z_i)\mathbf{W}_n]z_i^n \mid \mathbf{W}_n \leqslant -B\} = 1 \quad \text{for } j=1,2$$

Since n is the number of steps until absorption, it is certain that $\mathbf{W}_n \geqslant A$ or $\mathbf{W}_n \leqslant -B$. If the overshoot of the barriers is negligible, then $\mathbf{W}_n = A$ when response R_A is made and $\mathbf{W}_n = -B$ when R_B is given. These approximations reduce the above expression to

$$P(R_A \mid S_i)\exp[-\theta_j(z_i)A]E_A(z_i^n) + P(R_B \mid S_i)\exp[\theta_j(z_i)B]E_B(z_i^n) = 1 \quad \text{for } j=1,2$$

where, to simplify notation we have let $E_A(z_i^n) = E(z_i^n \mid \mathbf{W}_n \geqslant A)$ and $E_B(z_i^n) = E(z_i^n \mid \mathbf{W}_n \leqslant -B)$.

We now solve these two equations for

$$P(R_A \mid S_i)E_A(z_i^n) \quad \text{and for} \quad P(R_B \mid S_i)E_B(z_i^n)$$

giving us

$$P(R_A \mid S_i)E_A(z_i^n) = \frac{\exp[\theta_2(z_i)B]-\exp[\theta_1(z_i)B]}{\exp[\theta_2(z_i)B-\theta_1(z_i)A]-\exp[\theta_1(z_i)B-\theta_2(z_i)A]} \tag{10.2}$$

and

$$P(R_B \mid S_i)E_B(z_i^n) = \frac{\exp[-\theta_1(z_i)A]-\exp[-\theta_2(z_i)A]}{\exp[\theta_2(z_i)B-\theta_1(z_i)A]-\exp[\theta_1(z_i)B-\theta_2(z_i)A]} \tag{10.3}$$

Now, to derive the probability of an error when stimulus S_A is presented, $P(R_B \mid S_A)$, we need only let $i=A$ and set $z_A=1$ in Eq. 10.3. When $z_A=1$, $E_B(z_A^n)=1$, $\theta_1(z_A)=0$, and $\theta_2(z_A)=\theta_A$. Thus

$$P(R_B \mid S_A) = \frac{1-\exp(-\theta_A A)}{\exp(\theta_A B)-\exp(-\theta_A A)}$$

Similarly, to derive the predicted error probability when the stimulus is S_B we first set $i=B$ and then set $z_B=1$ in Eq. 10.2.

Derivation of the mean RTs is a little more complicated but still straightforward. Note first that under the appropriate regularity conditions:[1]

$$\frac{d}{dz_i}E_A(z_i^n)\Big|_{z_i=1} = E_A\left[\frac{d}{dz_i}(z_i^n)\right]\Big|_{z_i=1} = E_A(nz_i^{n-1})\Big|_{z_i=1} = \bar{T}_{iA}$$

Thus to estimate the conditional mean RTs we need only differentiate both sides of Eq. 10.2 and Eq. 10.3 with respect to z_i and then set z_i equal to 1.

When we perform this calculation we repeatedly run into the term $d[\theta_j(z_i)]/dz_i$, and thus we must evaluate it before we can conclude the derivation. Link and Heath (1975) suggest the following procedure. First note that because of the way we defined z_i it is always true that $z_i M_i(\theta)=1$. Therefore,

$$0 = \frac{d}{dz_i}\{z_i M_i[\theta_j(z_i)]\}$$

$$= M_i[\theta_j(z_i)] + z_i \frac{d\{M_i[\theta_j(z_i)]\}}{dz_i}$$

$$= M_i[\theta_j(z_i)] + z_i \frac{d\{M_i[\theta_j(z_i)]\}}{d\theta_j(z_i)}\frac{d\theta_j(z_i)}{dz_i}$$

[1] Remember that we are assuming that Δt, the time it takes to complete one sample, equals 1 msec. This assumption is merely for convenience. For instance, the value of Δt clearly will not affect the direction of any mean RT orderings. Without this assumption the following equation reads

$$\Delta t \frac{d}{dz_i}E_A(z_i^n)\Big|_{z_i=1} = T_{iA}$$

Solving for $d\theta_j(z_i)/dz_i$, we find our solution:

$$\frac{d\theta_j(z_i)}{dz_i} = \frac{-M_i[\theta_j(z_i)]}{z_i\langle d\{M_i[\theta_j(z_i)]\}/d\theta_j(z_i)\rangle}$$

To evaluate this derivative at $z_i=1$, recall that $\theta_1(1)=0$ and $\theta_2(1)=\theta_i$ (for $i=A$ or B) and that the derivative of the mgf at $\theta=0$ is $-\mu_i$ whereas the derivative at $\theta=\theta_i$ is $c_i\mu_i$. Thus

$$\left.\frac{d\theta_1(z_i)}{dz_i}\right|_{z_i=1} = \frac{1}{\mu_i} \quad \text{and} \quad \left.\frac{d\theta_2(z_i)}{dz_i}\right|_{z_i=1} = -\frac{1}{c_i\mu_i}$$

Using these results, the differentiation needed to estimate the conditional mean RTs is quickly completed, yielding the solutions. □

The integrity of these approximations depends on the magnitude of the overshoot when the walk first crosses a barrier. The derivations assume no overshoot occurs, and so if the average overshoot is small, the approximations will be good.

One way to get an indication of the magnitude of the overshoot, and thus about how good a specific set of approximations are, is to compare the estimated values of the average step sizes (i.e., the u_i) with the A and B estimates. If, say, $\hat{A}$ is large relative to $\hat{\mu}_A$, then the approximations should be fairly good; but if $\hat{A}$ is not substantially larger than $\hat{\mu}_a$, the overshoot may be significant enough to make the approximations poor.

This very general random walk model can predict any RT ordering. It has eight unconstrained parameters:[2] A, B, μ_A, μ_B, θ_A, θ_B, c_A, and c_B. In experiments of the type we have been considering, there are, at most, six degrees of freedom: four mean RTs and two error probabilities. It is therefore clear that to construct a testable random walk model the parameters of this general model must be constrained in some fashion. The SPRT model constrains the general model so that only four free parameters remain while the unbiasedness version of the RJT model constrains it so that five remain. Thus, both of these models are testable special cases of the general random walk model, but as we will see, the two models make markedly different mean RT predictions.

The SPRT model

Before we can use Proposition 10.1 to generate predictions of this model we must know something more about the six parameters, μ_A, μ_B, θ_A, θ_B, c_A, and c_B. Since there are only four free parameters in the model, some of these six must be functions of the others. These six parameters are all determined by the step size mgfs $M_A(\theta)$ and $M_B(\theta)$. Therefore, to utilize Proposition 10.1 it is necessary to examine the step size mgfs and to look for relationships between them.

[2] The parameters are unconstrained except for the obvious conditions that A, B, μ_A, θ_A, c_A, and c_B are positive and μ_B and θ_B are negative.

To begin, recall that

$$M_B(\theta) = \int_{-\infty}^{\infty} e^{-\theta x} g_B(x)\, dx$$

where in the SPRT model $g_B(x)$ is the probability density function of the random step size $\mathbf{X}=\ln[f_A(\mathbf{Y})/f_B(\mathbf{Y})]$.

This is, of course, just the expectation of $e^{-\theta \mathbf{X}}$ with respect to the density $g_B(x)$. It is well known that this expectation is identical to the one derived from $f_B(y)$ (the density on the random sensory value $\mathbf{Y}$ when the stimulus is S_B), when $\mathbf{X}$ is considered as a function of $\mathbf{Y}$; thus

$$\begin{aligned} M_B(\theta) &= \int_{-\infty}^{\infty} e^{-\theta x} g_B(x)\, dx \\ &= \int_{-\infty}^{\infty} \exp\left[-\theta \ln \frac{f_A(y)}{f_B(y)}\right] f_B(y)\, dy \end{aligned}$$

where we assume that both x and y range over $(-\infty, +\infty)$.

The next result uses this relationship to show how $M_A(\theta)$ and $M_B(\theta)$ are related in the SPRT model.

Proposition 10.2: In the SPRT random walk model the step size mgfs $M_A(\theta)$ and $M_B(\theta)$ are related by $M_B(\theta)=M_A(\theta+1)$.

Proof:

$$\begin{aligned} M_B(\theta) &= \int_{-\infty}^{\infty} \exp\left\{-\theta \ln\left[\frac{f_A(y)}{f_B(y)}\right]\right\} f_B(y)\, dy \\ &= \int_{-\infty}^{\infty} \left[\frac{f_A(y)}{f_B(y)}\right]^{-\theta} \left[\frac{f_B(y)}{f_A(y)}\right] f_A(y)\, dy \\ &= \int_{-\infty}^{\infty} \left[\frac{f_A(y)}{f_B(y)}\right]^{-\theta} \left[\frac{f_A(y)}{f_B(y)}\right]^{-1} f_A(y)\, dy \\ &= \int_{-\infty}^{\infty} \left[\frac{f_A(y)}{f_B(y)}\right]^{-(\theta+1)} f_A(y)\, dy \\ &= \int_{-\infty}^{\infty} \exp\left\{-(\theta+1) \ln\left[\frac{f_A(y)}{f_B(y)}\right]\right\} f_A(y)\, dy \\ &= M_A(\theta+1) \quad \square \end{aligned}$$

We see that the two step size mgfs of the SPRT model are related by a simple translation, and so if we know one mgf, the other can be recovered exactly.[3] This means that the parameters of $M_B(\theta)$ can all be written as func-

[3] In addition, it is easy to show that if $M_B(\theta)=M_A(\theta+1)$, then the two step size densities are related by $f_B(x)=e^x f_A(x)$.

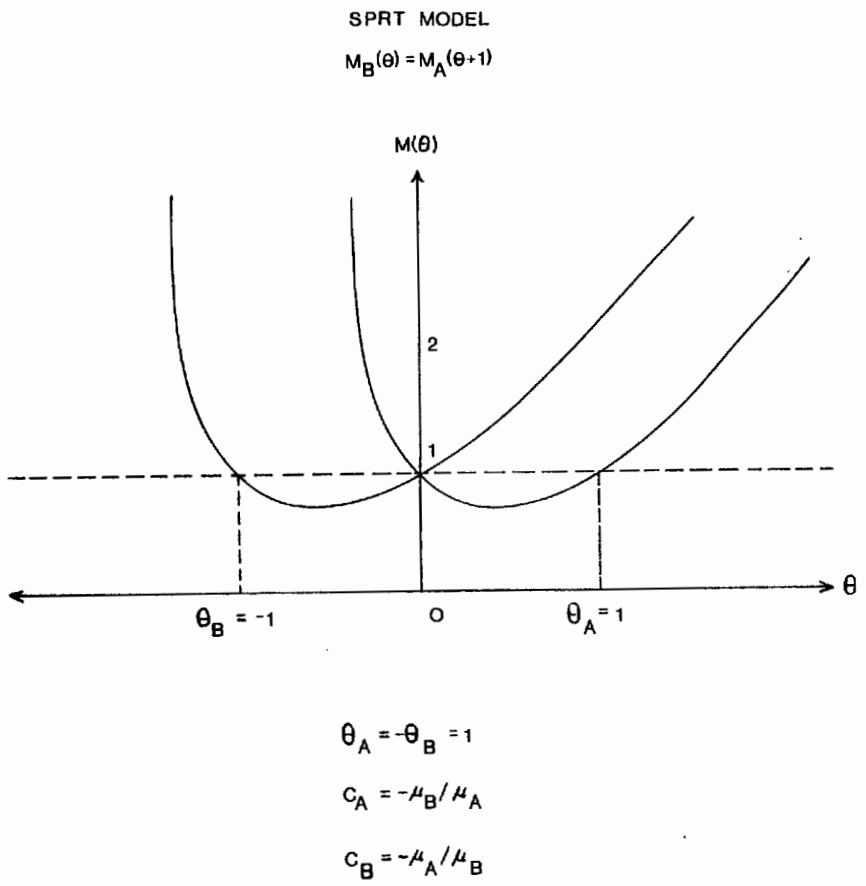


Fig. 10.4. An example of a pair of step size moment generating functions (mgfs) predicted by the SPRT random walk model. The two mgfs, which are related by a translation, satisfy the relation $M_B(\theta)=M_A(\theta+1)$.

tions of the $M_A(\theta)$ parameters and vice versa. In addition, we see immediately that when $\theta=0$, $M_B(0)=1$ and hence $M_A(1)=1$. Thus the second root of $M_A(\theta)=1$ is $\theta_A=1$. In a similar fashion, when we set $\theta=-1$, we see that $M_B(-1)=1$ and hence $\theta_B=-1$. In this way the model postulates four fewer parameters than the general model, leaving four.

An example of a pair of mgfs related by such a translation is illustrated in Fig. 10.4. Notice that the mgfs are not necessarily symmetric, although they may be. This point turns out to be important in the comparison of the SPRT model to the RJT model assuming unbiasedness. A moment's thought and perhaps a quick glance at Fig. 10.4 will convince one that the parameters of the two translated mgfs are related in the following manner: $-\mu_A=c_B\mu_B$ and $-\mu_B=c_A\mu_A$. Thus the $M_B(\theta)$ parameters can indeed be written as functions of the $M_A(\theta)$ parameters: $\mu_B=-c_A\mu_A$ and $c_B=1/c_A$, while $\theta_A=1$ and $\theta_B=-1$.

We can now use these mappings and Proposition 10.1 to estimate the error probabilities and the conditional mean RTs predicted by the SPRT model. First, for the error probabilities, we find the model results in the following.

Corollary 10.1: In the SPRT random walk model,

$$P(R_B \mid S_A)=\frac{1-e^{-A}}{e^{B}-e^{-A}} \quad \text{and} \quad P(R_A \mid S_B)=\frac{e^{-B}-1}{e^{-B}-e^{A}} \quad \square$$

The error probabilities are predicted to be equal if the two barriers are equidistant from the origin (i.e., if $A=B$). If the barriers are not equidistant, then the response associated with the closest barrier will be more frequently made and hence will be associated with a higher error rate. In any case, we can see that by adjusting the distances from the origin to the barriers any desired ordering on the response probabilities can be achieved.

Corollary 10.1 also makes clear that the SPRT response probabilities depend only on the distances to the barriers and not on the mgf parameters μ_A and c_A. This unexpected result suggests that a very easy way to estimate the barrier parameters is to solve the two equations for A and B. This strategy yields the estimates

$$A=\ln\left[\frac{P(R_A \mid S_A)}{P(R_A \mid S_B)}\right] \quad \text{and} \quad B=\ln\left[\frac{P(R_B \mid S_B)}{P(R_B \mid S_A)}\right]$$

The next result gives the mean RT predictions of the SPRT model.

Corollary 10.2: In the SPRT random walk model, the conditional mean RTs are

$$\bar{T}_{iA}=\frac{1}{c_A\mu_A}\left[\frac{(Ac_A+B)e^{B}+(A+Bc_A)e^{-A}}{e^{B}-e^{-A}}-\frac{Bc_A+Be^{B}}{e^{B}-1}\right]$$

and

$$\bar{T}_{iB}=\frac{1}{c_A\mu_A}\left[\frac{(Ac_A+B)e^{B}+(A+Bc_A)e^{-A}}{e^{B}-e^{-A}}+\frac{Ac_A+Ae^{-A}}{e^{-A}-1}\right]$$

where again i represents stimulus S_A or S_B. $\square$

The first and most important thing to note about these equations is that neither depends on the subscript i. Thus $\bar{T}_{AA}=\bar{T}_{BA}$ and $\bar{T}_{AB}=\bar{T}_{BB}$. In other words, mean RT, when conditioned on the response made, is predicted by the SPRT model to be the same for correct and incorrect responses. This is the prediction for which the model is, perhaps, most famous (or infamous). We saw in the preceding chapter that it is a prediction unlikely to be empirically supported and for this reason the simple SPRT model can be rejected as a complete description of human choice RT.

The RJT unbiasedness model

To generate predictions of the RJT model assuming unbiasedness, we must follow the same steps we did with the SPRT model. Our task is now easier, however, since the assumption of unbiasedness places constraints directly on the step size density functions rather than indirectly, as in the SPRT model. Specifically, the assumption manifests itself as $g_B(-x)=g_A(x)$, so that the densities are mirror images of each other. The mgfs are related in a similar manner. First note that

$$M_A(\theta)=\int_{-\infty}^{\infty} e^{-\theta x} g_A(x)\,dx=\int_{-\infty}^{\infty} e^{-\theta x} g_B(-x)\,dx$$

Now making the change of variable $u=-x$ [so $dx=-du$ and $g_B(u)=g_B(-x)$] yields

$$M_A(\theta)=-\int_{+\infty}^{-\infty} e^{\theta u} g_B(u)\,du=\int_{-\infty}^{+\infty} e^{\theta u} g_B(u)\,du=M_B(-\theta)$$

Thus, the mgfs are also mirror images. As in the SPRT model, if one of these mgfs is known, the other can be recovered exactly and so its parameters can always be written as a function of the parameters of the first mgf.

An example of two mgfs that are mirror images is given in Fig. 10.5. As was the case with the SPRT model, these mgfs need not be symmetric, although they might be. A casual examination of Fig. 10.5 reveals that the parameters of the general model must satisfy the following constraints when the step size mgfs are mirror images: $\mu_A=-\mu_B$, $\theta_A=-\theta_B$, and $c_A=c_B$. The model has five free parameters (A, B, μ_A, θ_A, and c_A) - one more than the SPRT model.

Using these equivalence mappings and Proposition 10.1, we can quickly derive predicted statistics. First the error probabilities:

Corollary 10.3: All RJT random walk models assuming unbiasedness predict that

$$P(R_B \mid S_A)=\frac{1-\exp(-\theta_A A)}{\exp(\theta_A B)-\exp(-\theta_A A)}$$

and

$$P(R_A \mid S_B)=\frac{\exp(-\theta_A B)-1}{\exp(-\theta_A B)-\exp(\theta_A A)} \quad \square$$

These equations are very similar to the SPRT predictions given in Corollary 10.1. In fact, if we set the SPRT parameter A equal to $\theta_A A$ in the RJT model and B equal to $\theta_A B$, the two models will give identical response probability predictions. Therefore, one cannot discriminate between them on the basis of accuracy data alone.

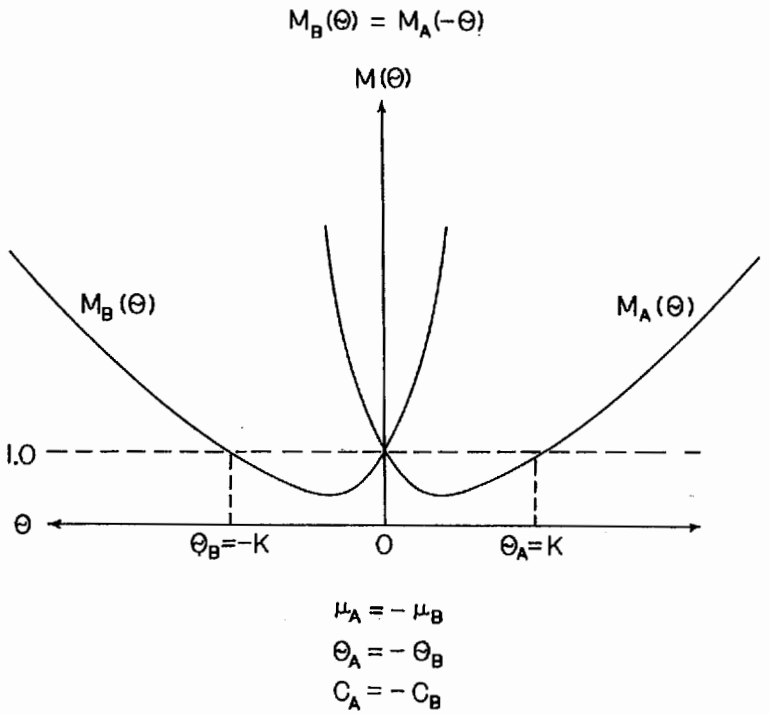


Fig. 10.5. An example of a pair of step size moment generating functions (mgfs) predicted by the RJT unbiasedness random walk model. The two mgfs, which are mirror images, satisfy the relation $M_B(\theta)=M_A(-\theta)$.

The latency predictions are a different story, however. Here the predictions of the two models sharply diverge. Link and Heath (1975) showed that the unbiasedness version of the RJT model can predict correct latencies to be greater than, less than, or equal to error latencies. Specifically, by algebraically manipulating the conditional mean RTs from Proposition 10.1 under the parameter constraints of the model, it is possible to prove the following.

Corollary 10.4: All RJT random walk models assuming unbiasedness predict that the difference in conditional mean RTs is

$$\bar{T}_{AA}-\bar{T}_{BA}=\frac{A}{\mu_A}\left(\frac{c_A-1}{c_A}\right)$$

and

$$\bar{T}_{BB}-\bar{T}_{AB}=\frac{B}{\mu_A}\left(\frac{c_A-1}{c_A}\right) \quad \square$$

If $c_A = 1$ and the mgfs are symmetric, these differences are zero; if $c_A > 1$, the differences are both positive and correct latencies are longer than error latencies. Finally, if $c_A < 1$, error RTs are longer than correct RTs. Thus the ordering must be the same for both responses, that is, either $\bar{T}_{AA} > \bar{T}_{BA}$ *and* $\bar{T}_{BB} > \bar{T}_{AB}$, or $\bar{T}_{AA} < \bar{T}_{BA}$ *and* $\bar{T}_{BB} < \bar{T}_{AB}$. We saw in the preceding chapter, however, that there are occasional reports in the literature of different orderings on different responses. For instance, both Snodgrass et al. (1967) and Carterette et al. (1965) found mean RT orderings of the form $\bar{T}_{AA} < \bar{T}_{BA} < \bar{T}_{AB} < \bar{T}_{BB}$, where the stimuli have been renamed to aid comparison (recall that $\bar{T}_{ij}$ is the mean RT given response R_j and stimulus S_i). Thus, when responding A, observers were faster when they were correct; but when responding B, incorrect responses were faster. This result can be predicted by neither the SPRT model or the unbiasedness version of the RJT model. In the next section, however, we will encounter a generalized version of the SPRT model that easily predicts orderings of this type.

Corollary 10.4 also reveals the importance of the parameters A and B in the predicted mean RT differences. More specifically, we can see that

$$\frac{\bar{T}_{AA} - \bar{T}_{BA}}{\bar{T}_{BB} - \bar{T}_{AB}} = \frac{A}{B}$$

Thus if $A = B$, the two differences are predicted to be equal, whereas if $A > B$, the difference between A response times will be greater than the difference between B response times.

As it turns out, we can also use Corollary 10.3 to estimate this same A/B ratio. The nice thing about this fact is that it provides a testable prediction of the RJT unbiasedness model that does not require the estimation of any parameters.

Proposition 10.3: All RJT random walk models assuming unbiasedness predict that

$$\frac{\ln P(R_A \mid S_A) - \ln P(R_A \mid S_B)}{\ln P(R_B \mid S_B) - \ln P(R_B \mid S_A)} = \frac{\bar{T}_{AA} - \bar{T}_{BA}}{\bar{T}_{BB} - \bar{T}_{AB}}$$

Proof: From Corollary 10.3,

$$\frac{\ln P(R_A \mid S_A) - \ln P(R_A \mid S_B)}{\ln P(R_B \mid S_B) - \ln P(R_B \mid S_A)} = \frac{\theta_A A}{\theta_A B} = \frac{A}{B}$$

whereas from Corollary 10.4 we saw that

$$\frac{\bar{T}_{AA} - \bar{T}_{BA}}{\bar{T}_{BB} - \bar{T}_{AB}} = \frac{A}{B} \quad \square$$

To get some rough idea as to whether or not this prediction is reasonable, let us apply it to the data from the Townsend and Snodgrass (1977) lateral masking study we discussed in the preceding chapter. This experiment provides two independent tests of the prediction. In one condition the stimuli to be recognized were the letters O and D and in a second condition the stimuli were the letters I and L. The estimated mean RTs and response probabilities are given in Tables 9.1 and 9.2.

If we substitute the appropriate estimates into the Proposition 10.3 test, we find that

$$\frac{\ln \hat{P}(R_O \mid S_O) - \ln \hat{P}(R_O \mid S_D)}{\ln \hat{P}(R_D \mid S_D) - \ln \hat{P}(R_D \mid S_O)} = .81 \quad \text{and} \quad \frac{\bar{T}_{OO} - \bar{T}_{DO}}{\bar{T}_{DD} - \bar{T}_{OD}} = .98$$

whereas

$$\frac{\ln \hat{P}(R_I \mid S_I) - \ln \hat{P}(R_I \mid S_L)}{\ln \hat{P}(R_L \mid S_L) - \ln \hat{P}(R_L \mid S_I)} = .82 \quad \text{and} \quad \frac{\bar{T}_{II} - \bar{T}_{LI}}{\bar{T}_{LL} - \bar{T}_{IL}} = .57$$

The correspondence is not perfect, although it must be remembered that there is probably a fairly large standard error associated with each of these statistics. One thing in the model's favor is that in each of the two cases both estimates of the A/B ratio are consistent about which barrier is farther from the origin. For the O, D stimulus pair, both the response probability and the mean RT estimates of A/B agree that the R_O barrier is closer to the origin than the R_D barrier. For the other pair of stimuli the two estimates agree that the R_I barrier is closer than the R_L barrier. This outcome concurs with the status of O and I as frequently employed vowels in the English language. While the application of this test to the Townsend and Snodgrass (1977) data is not conclusive enough to allow us to either accept or reject the RJT unbiasedness model, the agreement is close enough to that predicted by the model for us to at least conclude that the Proposition 10.3 prediction is not an unreasonable one to make.

We have now briefly examined two classes of tractable random walk models: those (of the SPRT) in which the step size mgfs are related by a unit translation and those (RJT) in which the mgfs are mirror images. Examples of both classes are illustrated in Fig. 10.6. There it is seen that neither of these two families contains the other; there are pairs of mgfs differing only by a unit translation that are not mirror images and vice versa. Thus although the SPRT model is a special case of the general model, it is not contained within the RJT "unbiasedness" model (and neither does it contain it). Figure 10.6 also makes clear that the intersection of these two families are those pairs of mgfs that are both symmetric and differ only by a unit shift along the θ axis.

It is in just these cases (i.e., when $c = 1$) that the RJT model, assuming unbiasedness, predicts equivalent correct and error latencies (i.e., when conditioned on a response) that we have seen are characteristic predictions of the SPRT model. In fact, Swensson and Green (1977) showed that this prediction holds if and only if the step size mgfs are related by a shift.

Thus, it is the fact that the two mgfs of the SPRT model are always related by a translation that restricts the model's generality on this prediction. Since

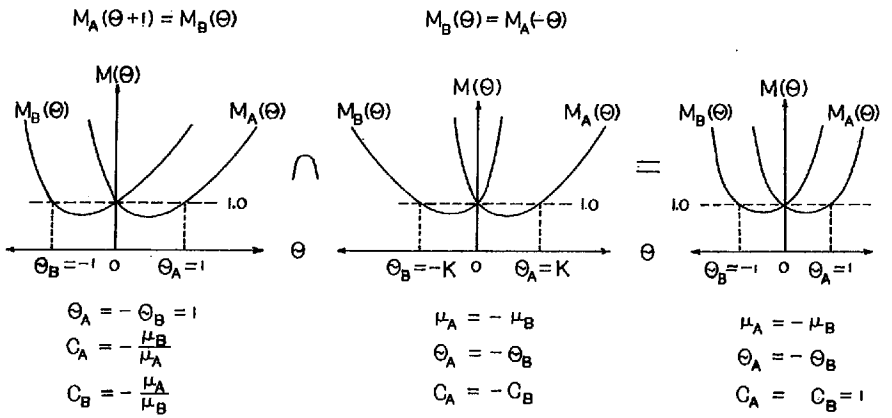


Fig. 10.6. Pairs of step size moment generating functions (mgfs). The left-most pair is an example of the translated mgfs predicted by the SPRT model. The central pair is an example of the mirror-image mgfs predicted by the RJT unbiasedness model and the right-most pair is an example of mgfs that can be predicted by both models.

some but not all mgfs that are mirror images are also simple translations, the unbiasedness version of the RJT model is more general on this prediction.

On the other hand, there are pairs of mgfs fulfilling $M_A(\theta+1)=M_B(\theta)$ but not $M_A(-\theta)=M_B(\theta)$. Figure 10.6 contains one such example. This leads us to wonder if there are some predictions made by the SPRT model that cannot be made by the RJT model assuming unbiasedness. Just what these predictions might be is difficult to say. We saw earlier that the models make almost identical predictions about response probability, so any difference between the models must be in how they predict mean RT, especially when it is conditioned on the stimulus presented. Expressions for the mean conditioned response times are very cumbersome and difficult to manipulate and, to our knowledge, no such divergent predictions have yet been found.

More general random walk models

Generalizing the SPRT model

The primary assumption of the SPRT model is that the observer transforms the original psychological continuum by the mapping $\mathbf{Y}\rightarrow \ln[f_A(\mathbf{Y})/f_B(\mathbf{Y})]$. This mapping is completely observer-controlled. Wald and Wolfowitz (1948) showed that if the barriers are equally distant from the start point, the mapping is optimal in the sense that for any given error probabilities the mean response time is minimized. The model assumes that the observer has discovered this optimality through experience. Unfortunately,

the data disagree. Observers apparently do not always apply this transformation to the psychological axis. This is not to say, though, that they do not apply it sometimes or that they do not apply some other transformation. If we generalize the SPRT model by assuming only that *some* transformation of the psychological axis is made (one possible example of which is $\mathbf{Y}\rightarrow \ln[f_A(\mathbf{Y})/f_B(\mathbf{Y})]$), the model becomes much more general and, like the most general RJT model, much more difficult to refute. In fact, the subjectivity of this transformation provides the defender of this general model an almost irrefutable position: Observers may adopt different transformations in different experimental situations. Falsification is now very difficult if not outright impossible.

The problem, however, is to come up with a viable alternative transformation. One possibility, recently suggested by Ashby (1983), is to introduce more potential for bias into the discrimination process. In virtually all extant random walk models the only structure available for introducing bias into the system is in the relative placement of the barriers at A and $-B$.[4] In such models, it is assumed that the accumulation process itself is conducted without bias. The bias enters only when the discrimination process terminates with different criterion levels of evidence for the different responses. It is not inconceivable to imagine a biased accumulation process in which evidence is misconstrued somewhat to favor one response more than it would if no bias existed. Thus, if an observer is biased toward response R_A, then instead of (or in addition to) decreasing the criterion for R_A responses, a biased accumulation model might predict, for instance, that the observer views ambivalent evidence as slightly favoring response R_A.

Ashby suggested that this type of a biasing mechanism could be introduced into the SPRT model by assuming that when an observer calculates the likelihood ratio of a given psychological sample, he or she multiplies the resulting value by a biasing constant k. Thus, instead of transforming the observed value $\mathbf{Y}$ into $\ln[f_A(\mathbf{Y})/f_B(\mathbf{Y})]$, this model assumes the observer performs the transformation

$$\mathbf{Y}\rightarrow \ln\left\{k\left[\frac{f_A(\mathbf{Y})}{f_B(\mathbf{Y})}\right]\right\}$$

A bias toward response R_A occurs if $k>1$, and a bias toward R_B occurs if $k<1$. If $k=1$, there is no bias in the accumulation process and the model reduces to the classical SPRT model. Thus the biased accumulation model clearly contains the SPRT model as a special case.

To obtain predictions of the model we need to examine its step size mgfs. First recall that

$$M_B(\theta)=\int_{-\infty}^{\infty} e^{-\theta x} g_B(x)\,dx$$

[4] Another somewhat similar possibility is that the walk might not begin at the origin (Link 1975, 1978).

where $\mathbf{X}=\ln[k[f_A(\mathbf{Y})/f_B(\mathbf{Y})]]$ in this case. Following our earlier development for SPRT models, Ashby showed that in the biased SPRT model the step size mgfs are related as follows.

Proposition 10.4: In the biased version of the SPRT random walk model, the step size mgfs $M_A(\theta)$ and $M_B(\theta)$ are related by $M_B(\theta)=kM_A(\theta+1)$.

Proof:

$$
\begin{aligned}
M_B(\theta) &= \int_{-\infty}^{\infty} \exp\left\{-\theta \ln\left[k\frac{f_A(y)}{f_B(y)}\right]\right\} f_B(y)\,dy \\
&= \int_{-\infty}^{\infty} \exp\left\{\ln\left[k\frac{f_A(y)}{f_B(y)}\right]^{-\theta}\right\} f_B(y)\,dy \\
&= \int_{-\infty}^{\infty} k^{-\theta}\left[\frac{f_A(y)}{f_B(y)}\right]^{-\theta} f_B(y)\,dy \\
&= k\int_{-\infty}^{\infty} k^{-(\theta+1)}\left[\frac{f_A(y)}{f_B(y)}\right]^{-(\theta+1)}\left[\frac{f_A(y)}{f_B(y)}\right] f_B(y)\,dy \\
&= k\int_{-\infty}^{\infty} \exp\left\{\ln\left[k\frac{f_A(y)}{f_B(y)}\right]^{-(\theta+1)}\right\} f_A(y)\,dy \\
&= k\int_{-\infty}^{\infty} \exp\left\{-(\theta+1)\ln\left[k\frac{f_A(y)}{f_B(y)}\right]\right\} f_A(y)\,dy \\
&= kM_A(\theta+1) \quad \square
\end{aligned}
$$

Thus the two mgfs are not only related by a shift but in addition one is a multiple of the other.

Two examples of such mgfs are given in Fig. 10.7. The top panel illustrates a bias toward response R_A (i.e., $k>1$), and the bottom panel illustrates a bias toward R_B (i.e., $k<1$). It can be immediately seen that the two mgfs are not simple translations of one another. Translations will result only in the unbiased case of $k=1$ (i.e., the traditional SPRT model). This is important because it tells us that the biased SPRT model does not, in general, make the unfortunate prediction that correct and incorrect mean RTs must always be equal (when conditioned on a specific response). The model must thus be viewed as a more serious alternative to the RJT unbiasedness model than is the classical SPRT model.

In addition, the model turns out to have a very natural and appealing interpretation of the biasing constant k. Suppose that instead of basing the decision process on the conditional densities $f(y\mid S_i)$, the observer instead utilizes the a posteriori probabilities $f(S_i\mid y)$. The latter seems the more natural choice of a real processing system. Given the sensory value $\mathbf{Y}$, the system

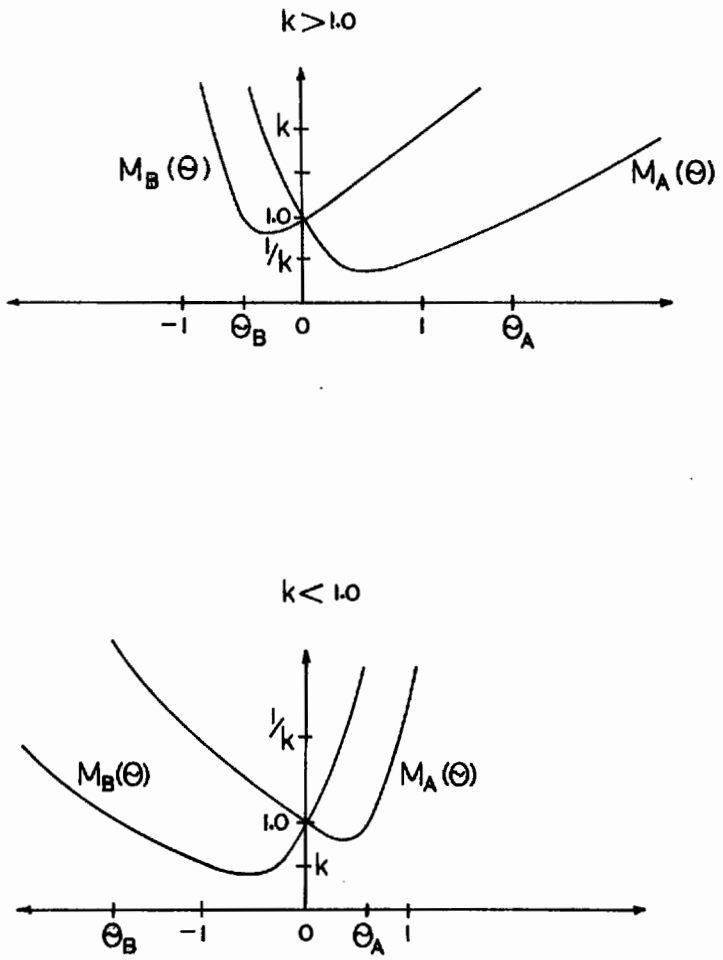


Fig. 10.7. Examples of step size moment generating functions (mgfs) predicted by the biased SPRT model. The mgfs are related by $M_B(\theta)=kM_A(\theta+1)$. (a) represents a bias toward response R_A and (b) represents a bias toward R_B.

computes the probability that stimulus S_A was presented and compares this value with the probability that S_B was presented (via a likelihood ratio).

If the posterior probabilities are chosen, then the jth step size becomes

$$\mathbf{X}_j=\ln\left[\frac{f(S_A\mid\mathbf{Y}_j)}{f(S_B\mid\mathbf{Y}_j)}\right]$$

Using Bayes's theorem, this expression can be rewritten as

$$\mathbf{X}_j = \ln\left[\frac{f(\mathbf{Y}_j \mid S_A)P(S_A)/f(\mathbf{Y}_j)}{f(\mathbf{Y}_j \mid S_B)P(S_B)/f(\mathbf{Y}_j)}\right]$$

$$= \ln\left[\frac{P(S_A)}{P(S_B)}\,\frac{f(\mathbf{Y}_j \mid S_A)}{f(\mathbf{Y}_j \mid S_B)}\right]$$

and the biasing constant k can be interpreted as the ratio of the a priori stimulus presentation probabilities, that is, $k = P(S_A)/P(S_B)$. Thus this interpretation predicts that no bias exists when $P(S_A) = P(S_B)$. When stimulus S_A is more likely to be presented there is a bias toward response R_A, and when S_B is more likely there is a bias toward R_B.

Under this interpretation, different kinds of bias may be produced when stimulus and when response parameters are manipulated. Response parameters (e.g., payoffs) may be more likely to affect barrier placements (i.e., the values of A and B), whereas stimulus parameters (e.g., stimulus probability) may sometimes affect the accumulation process itself. There seems to be no good reason, at this point at least, to expect the same sorts of bias to result from such different experimental manipulations.

One advantage this model has over the RJT unbiasedness model is that it can predict troublesome mean RT orderings of the form $\bar{T}_{AA} < \bar{T}_{BA} < \bar{T}_{AB} < \bar{T}_{BB}$ such as found by Snodgrass et al. (1967) and Carterette et al. (1965). In this and the preceding chapter we consistently found this type of RT ordering to be very difficult or impossible for most choice RT models to predict. One factor both studies had in common was that stimulus presentation probability was one of the independent variables manipulated. It may be that so many models have trouble with these orderings because, in such cases, a fundamentally different kind of bias is in effect, a bias in the accumulation process itself.

Generalizing the RJT model

Link (1978) recently proposed a method of testing a more general version of the RJT random walk model in which the constraint on the parameters is that $A = B$, so that the two barriers are equidistant from the origin. But, no restrictions are placed on the step size mgfs. They need not be related at all. This class of models is thus very large. Even so, it does not contain the unbiasedness model as a special case, since in many of these $A \neq B$.

Link based his test on the general prediction of all random walk models that, for a given stimulus S_i, when we average over responses, the expected number of steps until absorption equals the average distance to the barriers divided by the expected drift rate, that is,

$$E_i(n) = \frac{AP(R_A \mid S_i) - BP(R_B \mid S_i)}{\mu_i}$$

$$= \frac{AP(R_A \mid S_i) - B[1 - P(R_A \mid S_i)]}{\mu_i} =$$

$$= \frac{(A+B)P(R_A \mid S_i) - B}{\mu_i}$$

Now if $A = B$, this expression reduces to

$$E_i(n) = \frac{A}{\mu_i}\,[2P(R_A \mid S_i) - 1]$$

If we let t_0 equal the mean base time, then all random walk models for which $A = B$ predict that the mean RT when stimulus S_i is presented is

$$\bar{T}_i = \frac{A}{\mu_i}\,[2P(R_A \mid S_i) - 1] + t_0$$

This is a testable prediction. If, in some manner, we can induce observers to trade speed for accuracy across several experimental conditions, then mean RT to stimulus S_i should be a linear function of the observable statistic $[2P(R_A \mid S_i) - 1]$ with slope A/μ_i and intercept t_0. Any deviation from linearity falsifies this class of random walk models.

It is important to realize, however, that the SPRT model (or the biased SPRT model), with $A = B$, makes this same prediction. Thus the test is not one of SPRT models vs. RJT models but instead is *meant* to be a test of random walk models vs., say, counting models, although it has not yet been shown that counting models cannot also predict linear or near-linear functions in this situation. Because of this, if the prediction is supported in an empirical test, counting models (or any other possibilities) cannot be ruled out. On the other hand, if the prediction fails, then all random walk models with $A = B$ can be ruled out.

At any rate, Link (1978) tested this prediction with data from three different experiments and found that in every case the T_i vs. $[2P(R_A \mid S_i) - 1]$ plots were remarkably linear. A large class of random walk models thus receives tentative support.

Conclusions

Random walk models are powerful models of human choice RT and accuracy. In general, these models tend to be very versatile in the types of mean RT orderings they can predict, much more so than, say, the counting models we considered in the preceding chapter. Several disadvantages accompany this versatility, however. One is that we can only approximate the mean RTs predicted by random walk models, and in some cases these approximations may not be very good. In general, they will only be good when the overshoot of the response barriers is negligible.

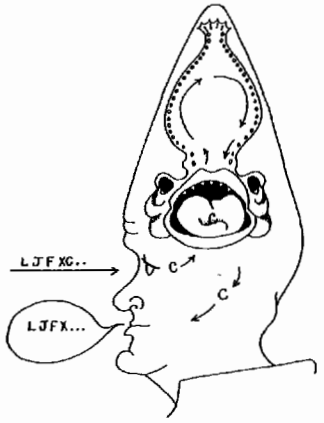
A second disadvantage associated with random walk models is that, even when the approximations are good, the mean RT predictions (from Proposition 10.1) are cumbersome and difficult to manipulate. This makes it very hard in general to determine just what RT orderings an arbitrary random walk model can predict.

We focused our discussion in this chapter on two different random walk models, the SPRT model developed by Stone (1960) and elaborated by Laming (1968) and the RJT model assuming unbiasedness of Link and Heath (1975). In addition to yielding quite different predictions, these models also aim their postulates at different levels of processing. The restriction that the step size mgfs of the SPRT model be translations of one another results entirely from the nature of the transformation that is assumed to be made on the sampled sensory values. No assumptions are ever made about the distribution of these sensory values. On the other hand, in the RJT unbiasedness model, the restriction that the step size mgfs be mirror images does place restrictions on the distributions of the sensory values. Thus RJT models make assumptions about sensory events, whereas SPRT type models make assumptions primarily about decision strategies.

These models should be testable on this decision vs. sensory basis. For instance, an experimental manipulation of the parameters of the physical stimuli (e.g., intensity) should affect the psychological stimulus distributions, but by itself should not affect the way an observer uses these distributions in his or her decision process. Thus if an observer calculates log likelihood ratios in a given experimental situation and then a second experiment is conducted under exactly the same conditions as the first, except that, say, stimulus intensity is changed somewhat, it is reasonable to expect the observer to again calculate log likelihood ratios. In other words, if the SPRT model is found to accurately predict a given set of experimental results, it should also be able to predict results from an experimental situation in which only stimulus changes have been made.

On the other hand, manipulation of response parameters (e.g., payoffs etc.) may cause an observer to begin calculating something other than strict log likelihood ratios, but it should not affect the psychological stimulus distributions. There should thus be no reason for the observer to alter his internal referent in such situations, if he uses one in his decision process. This means that across-experiment changes in *response* parameters should not adversely affect the fit of RJT models.

In the next chapter we will conclude our examination of models that make predictions about accuracy as well as latency. These models will turn out to be very different from those we have discussed in this and the preceding chapter. The differences are due mainly to the very different response instructions given to observers. Instead of asking the observer to search for the presence or absence of some critical item or to identify one of two possible stimulus alternatives, in whole report paradigms the observer is asked to recall an entire multisymbol stimulus array. As we shall see, the modeling of whole report data involves some quite different yet interesting theoretical problems.



A *visual* whole report task. The letters are first input through the peripheral visual system. Inside Mr. Conehead's cranium lives a helpful homunculous who identifies symbols by eating them. The names of the symbols may then be reported on output by the verbal report system. The last letter in the stimulus list, *C*, is now being processed and prepared for verbal report.

11 *Investigating the processing characteristics of visual whole report behavior*

In a *whole-report* task, the observer tries to report back all the symbols in the stimulus display. Partial report (Sperling 1960) and detection procedures (Estes & Taylor 1964), however, permit the observer to report a limited amount of information. The partial report technique designates only a portion of the display for a report, whereas the detection procedure requires the observer to report which of two signal symbols was placed in an array of distractor ("noise") symbols.

These latter designs effectively lessen the immediate memory load, thus allowing the investigation of limitations on other stages in the information-processing chain. The basic partial report procedure permits a delimitation of the quantity of information initially available, whereas the detection procedure yields an estimate of the average number of display symbols compared with two memory symbols. A wide variety of techniques in addition to these have added to our understanding of visual cognitive processes.

Whole report continues to be of interest, however. The mechanisms contributing to whole report performance are not yet entirely delineated. It has even remained a question as to where in the processing chain the major losses of information occur, although recent suggestive evidence has surfaced that will be considered below. In addition, the possibility was put forth recently

An abbreviated report of the study contained in this chapter appears in Townsend (1981).

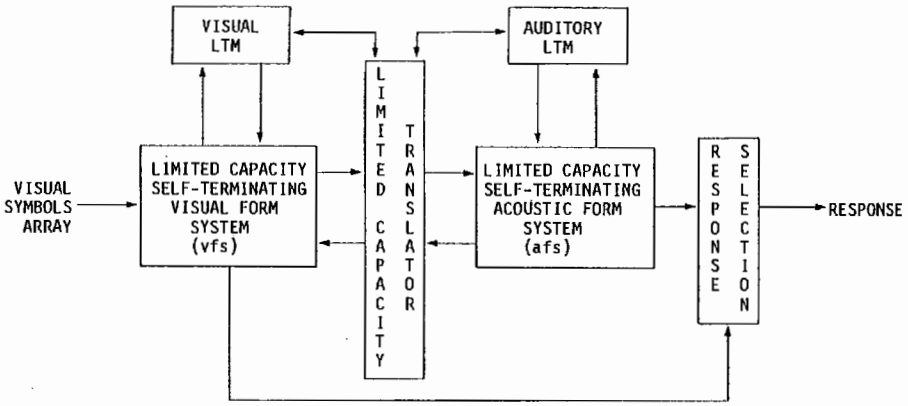


Fig. 11.1. A schema of selected subprocesses in visual processing. From Townsend and Roos (1973).

that partial report may not always be simpler than whole report in the processing steps called for (Scheerer 1974).

In the present chapter, we will consider some whole report data from the perspective offered by the approach and issues suggested in other parts of the book. In addition, some other recent findings that appear germane to a plausible model of the whole report processes will be discussed. It will aid our exposition to have a flowchart of the major postulated stages in whole report processing.

The schematic in Fig. 11.1 shows a flexible system that summarizes some of the functions that have been suggested to occur in a variety of visual symbol processing tasks. It includes the visual form system (vfs), which contains a subsystem responsible for iconic storage but may also be used for matching features against incoming visual information. A translating mechanism that utilizes information garnered from correlating visual long-term memory (LTM) identification with acoustic-verbal LTM to provide input to the acoustic form system (afs) follows. The translator may act in initiating a set of rehearsal instructions (Sperling 1967). The afs includes traditional immediate memory functions but also auditory icons and possible feature-altering propensities analogous to those of the vfs; acoustic inputs are not shown since the primary emphasis here is visual. The precise characteristics of the mechanisms responsible for these functions – for example, capacity limitations – are more controversial.

Whole report, by its very nature, must engage (1) complete identification of the symbols (vfs–visual LTM); and (2) a linkage, possibly independent of the visual identification, with stored acoustic-verbal information and some type of storage of the resulting acoustic-verbal form of the input (afs), before report. The typical partial report procedure (Averbach & Sperling 1961) also involves these functions, but does not require that the observer process as

much as possible from the display, and translation and afs storage requirements are minimal. The detection paradigm (Estes & Taylor 1964), on the other hand, may involve high usage of the vfs by permitting a storage of one or more target symbols or critical features in the vfs to be matched against as much of the incoming visual information as possible, although matching may terminate (self-termination) when the target is located. Subsequent translation to the afs of the recognized target then takes place.

Because of the distinct task demands of the various paradigms, the partial report and detection procedures do not establish true bounds on the capacity required to completely identify as many visually displayed symbols as possible; and the whole report task suffers the confounding of a high afs load. Later we will review some recent evidence concerning this type of vfs capacity and compare it with some suggestive information from the present experiment.

Another important question concerning the early stages of whole report processing (i.e., visual identification and translation) is whether they are serial or parallel. Sperling's work has been influential with regard to this issue (1963). He employed a whole report technique in conjunction with a noisy poststimulus mask with zero ISI (interstimulus interval) and plotted the average-number-correct curve as display interval varied. The results suggested a 10 msec per item scanning rate. This seems to have been widely interpreted at the time as evidence for serial processing, due to the linearity of the initial part of the curve.

However, some results reported later by Sperling (1967) appeared to suggest parallelity when serial position of the displayed letters was treated as a parameter in plotting the results. The reason is that the serial position curves were all gradually increasing, fanning out from a common origin with a negatively accelerated ascent.

Phase 1 of the experiment discussed below involved the acquisition of the joint probability correct information for the different serial positions (1–5) under conditions analogous to those used by Sperling (1967). The average (or marginal) probability correct for the distinct serial positions sweep out serial position functions when plotted against display duration. In addition to comparing these with Sperling's curves, the range of display times was extended and new analyses on the joint position information were performed. The display times were extended to 250 msec, about the maximum for which one could be reasonably assured that no eye movements occur. Interpretation of such serial position functions will be discussed below.

The new analyses of primary interest in Phase 1 concerned the question of whether the separate letters were processed independently of one another or whether conditioning on one or more other letters being correctly processed might raise or lower the likelihood that another was correctly processed (Townsend 1974a; Estes & Taylor 1966). Basically, if the information lost in whole report is due to a fixed-size short-term memory buffer (or, for that matter, any fixed-size channel or storage unit), then one would expect a nega-

tive correlation among the various stimulus positions. If, on the other hand, the first information loss occurs due to less than perfectly accurate but independent visual channels, then the accuracy probabilities would be predicted to be independent. Independence of total completion times (Chapter 4) implies the present type of independence. Finally, some models predict a positive correlation because of attentional fluctuations or the stochastic nature of the processing; a serial Poisson model is of the latter type. In fact, parallel models seem typically more intuitively compatible with independence, although serial models can, as we have seen in earlier chapters, predict independence as well. Mathematical models based on these three assumptions will be developed here and imbued with guessing structure. Guessing itself produces a negative correlation and certainly occurred in the experiment to be reported since a fixed number of responses were required on each trial. These models will then be tested against the dependence statistics from Phase 1.

Finally, the accuracy results from Phase 1 aided in the selection of display intervals used in Phase 2. Phase 2 of the experiment involved an interdisplay temporal manipulation to provide further information about spatiotemporal characteristics of processing. Eriksen and Spencer (1969) and Shiffrin and Gardner (1972) compared the effect of presenting one symbol or part of the display for t msec followed by another symbol for t msec and so on until the entire display had been shown (with k symbols, this results in a total display interval of $k \times t$ msec) with the effect of displaying the entire array simultaneously for only t msec. The former experiment (Eriksen & Spencer 1969) used a variety of sequential exposure times, the shortest being 5 msec, and therefore was effectively simultaneous. The Shiffrin and Gardner experiment (1972) utilized the same paradigm but with $t = 40$ msec. Surprisingly, the simultaneous condition produced accuracy about equal to that of the sequential condition. This result, of course, provides some intuitive support for parallel processing as opposed to serial processing in that specific paradigm. Both studies were of the recognition (or detection à la Estes & Taylor 1964) variety, and so are not directly pertinent for whole report processing. In any case, Townsend (1972; 169–92) discussed this type of paradigm with respect to parallel vs. serial processing and proposed an alternate design of similar spirit.

The idea in the new paradigm just reverses the Eriksen and Spencer (1969) logic. If each of k symbols is presented for t msec in the *sequential* condition, then the *simultaneous* presentation is set to last for $k \times t$ msec. The modified version of this design applied in Phase 2 employed a shield on the left or right part of the linear five-letter display during the early or late temporal fraction of the display duration respectively. Figure 11.2 illustrates the Phase 2 paradigm.

In this paradigm, if processing is parallel, one expects intuitively that the sequential type of presentation will degrade performance relative to the simultaneous, whereas if processing is serial and in an appropriate order, then little or no decrement in accuracy should be observed.

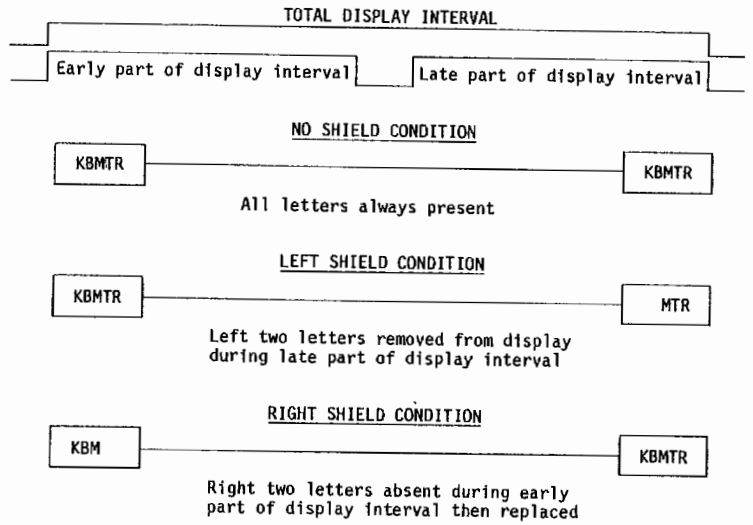


Fig. 11.2. A diagram of the Phase 2 paradigm.

Before taking up the details of the experimental design, it might be wise to consider some pertinent issues. The first concerns the "fanning-out" serial position functions mentioned above in connection with Sperling's (1967) results. We will show that even nonequivalent parallel and serial models can both produce behavior of this nature.

Secondly we want to indicate how the parallel or serial nature of a process can interact with the independence or dependence in affecting a type of statistic designed to test for seriality. Finally, an indication will be given that although stochastic serial models can predict some decrement in designs of the Phase 2 type, parallel models may produce more deterioration of performance.

Serial position curves and parallel vs. serial processing

Sperling (1967) provisionally rules out random-order serial processing (i.e., more than one processing order through the serial positions is taken from trial to trial) on the basis of his serial position curves for two reasons. Probably the most germane is that the pattern of results is "highly repeatable" from session to session. This might refer to (1) low variance in the processing times themselves, or (2) low variability in the observers' strategies across experimental sessions. The first is of little aid to the parallel hypothesis because it has been shown that independent parallel models sometimes predict larger variances than serial models (Townsend 1972: 191). The second is also not much help in this context because whether serial or parallel, an

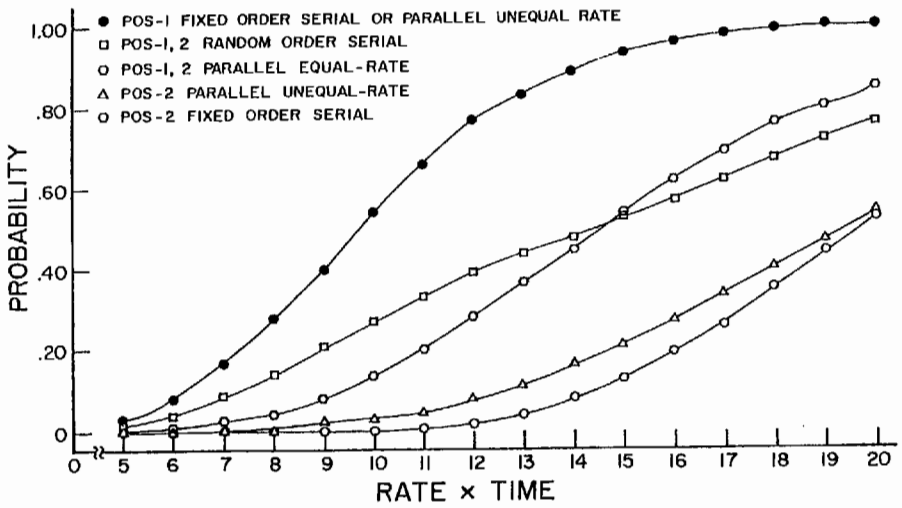


Fig. 11.3. Serial position curves from serial and parallel models that do predict serial position effects (bottom and top curves) or do not (middle two curves) for a two-item display.

observer may by dint of choice or inability to do otherwise process displays in a similar fashion from session to session.

In order to show that even quite different parallel and serial models can generate similar serial position curves, let us compare the serial position predictions of an independent parallel model to a garden variety serial model. By *independent* we refer to the same concept mentioned in connection with analyses of Phase 1 data above. Namely, independence requires that the completion of one item has absolutely no bearing upon the likelihood that the other has been completed. More precisely, this is the concept of independence of the total completion time of an element developed in Chapter 4. In other words, if $P(C_2 \mid C_1)$ is the probability correct on the letter in position 2 (by actual processing, not guessing), given that the processing system is correct on the letter in serial position 1, then $P(C_2 \mid C_1) = P(C_2)$ is the overall average probability correct on serial position 2, in the case of the independent parallel model. The standard-type serial model employed here predicts that $P(C_2 \mid C_1) > P(C_2)$.

To get some idea of the effects of search order and processing rate, the likelihood of different orders of processing (i.e., is the letter in position 1 processed first or is that in position 2 processed first?) was varied in the serial model and the rates of processing the letters in positions 1 and 2 were varied in the parallel model. The predictions of all four models are shown in Fig. 11.3. Both the serial and parallel models are based on underlying Poisson distributions; note that processing rate and duration combine multiplicatively. For instance, if rate $= u = 1$ item per 10 msec, then time in Fig. 11.3 would run from 50 to 200 msec. To increase the possibility of differential predictions, the number of stages (which might, for example, be associated with the number of features in letter stimuli) in the Poisson distributions was constrained to be the same in each hypothetical letter in the parallel and serial models (that is, number of stages $= 10$).

Two types of predictions are considered: those with strong serial position effects in which the probability correct on position 1 is different than it is on position 2; and those with no serial position effects, in which the probability correct is the same for all positions. The top curve gives the prediction for the fixed-order serial model on serial position 1 (probability of processing position 1 first $= 1$), and the bottom curve (open circles) gives its predictions on position 2. In the latter case, the serial function is, of course, just a Poisson function with twice the number of stages (20) as position 1 (10), because position 2 is always processed second and we assume that each letter contains the same number of features on the average.

Specifically, the fixed-order serial position 1 curve is the gamma distribution function

$$P_1(\mathbf{T}_1 \leqslant t) = \sum_{j=10}^{\infty} \frac{(ut)^j e^{-ut}}{j!}, \quad u = 1$$

whereas the position 2 curve is specified by

$$P_2(\mathbf{T}_2 \leqslant t) = \sum_{j=20}^{\infty} \frac{(ut)^j e^{-ut}}{j!}, \quad u = 1$$

These are also the respective probabilities of completing at least 10 or at least 20 stages, the presumed number of features processed in a Poisson counting mechanism. The parameter u equals 1 in Fig. 11.3. Strictly speaking, this and the other "models" discussed in this chapter are *classes* of models where each numerical specification of the parameter set yields a particular model. However, it is convenient here and in certain other parts of the book to refer to an appropriate general function of parameters (as variables) as the *model*.

The middle serial curve is the random-order model that assumes the probability of processing position 1 first is just $\frac{1}{2}$, with the attendant formula,

$$P_1(\mathbf{T}_1 \leqslant t) = P_2(\mathbf{T}_2 \leqslant t) = \frac{1}{2}\left[\sum_{j=10}^{\infty} \frac{(ut)^j e^{-ut}}{j!} + \sum_{j=20}^{\infty} \frac{(ut)^j e^{-ut}}{j!} \right], \quad u = 1$$

This model predicts no serial position effects.

The parallel model with serial position effects was gamma with the position 1 rate v_1 set to $v = u = 1$, and the numbers of stages to $k = 10$, so that the position 1 curve is identical to that of the fixed-order serial model (top curve in Fig. 11.3). The bottom parallel curve is for position 2 where v_2 was set to $v_1/2$, that is, one-half the speed of the position 1 processing rate. It was not manipulated to make it mimic the position 2 curve of the fixed-order serial model. The two parallel model curves are respectively predicted by

$$P_1(\mathbf{T}_1 \leqslant t) = \sum_{j=10}^{\infty} \frac{(vt)^j e^{-vt}}{j!}$$

$$P_2(\mathbf{T}_2 \leqslant t) = \sum_{j=10}^{\infty} \frac{(vt/2)^j e^{-(vt/2)}}{j!}, \quad v=1$$

Finally, the parallel curve that reflects the absence of serial position effects is shown in the middle of Fig. 11.3, and was obtained by arbitrarily letting $v_1=v_2=\frac{2}{3}u=\frac{2}{3}v=\frac{2}{3}(u=v=1)$ in the formula

$$P_i(\mathbf{T}_2 \leqslant t) = \sum_{j=10}^{\infty} \frac{(2vt/3)^j e^{-(2vt/3)}}{j!}, \quad i=1,2, \quad v=\tfrac{2}{3}$$

These parallel and serial models are really quite unalike in conception, differing on both the parallel–serial dimension and on the independent–dependent dimension, and with the abovementioned constraints acting to increase diversity in prediction. Yet the qualitative behavior of the serial and parallel functions are rather similar, with each revealing the fanning-out behavior usually associated with parallel models. Naturally, a serial model that is identical to the independent parallel model would generate completely equivalent serial position curves.

The independence question and a suggested method of testing for seriality

The issue of independence also arises in connection with another closely related method for testing between serial and parallel processing that was promulgated by Glezer and Nevskaya (1964). The idea of the suggested method is to plot the conditional probability of being correct on the second of two elements, given that the first was correct, $P(C_2 \mid C_1)$, as a function of the probability of being correct on the first $P(C_1)$ (with obvious generalizations to high numbers of elements). It is supposed that this function will, in the event that processing is fixed-order serial, ascend quite slowly until $P(C_1)$ is close to 1 and then rapidly accelerate, and there is some evidence for this trend in their data.

However, such a means of data presentation in some instances actually reduces parallel–serial discriminability. Consider again

$$P(C_2 \mid C_1) = \frac{P(C_1 \cap C_2)}{P(C_1)}$$

If the fact that the first element has completed processing increases the probability that the second has also been completed, as in the serial model above, then $P(C_2 \mid C_1) > P(C_2)$. But in the case of independent parallel processing, $P(C_2 \mid C_1) = P(C_2)$. That is, the serial conditional probability is elevated over the marginal $[P(C_2)]$, but the parallel conditional is just equal to the marginal. Figure 11.3 could in fact be replotted in the Glezer and Nevskaya manner. The parallel unequal-rate prediction (position 2) curve is simply rescaled to plot against $P(C_1)$, since $P(C_2 \mid C_1) = P(C_2)$. In the fixed-order serial model, $P(C_1 \cap C_2) = P(C_2)$ because for position 2 to have been

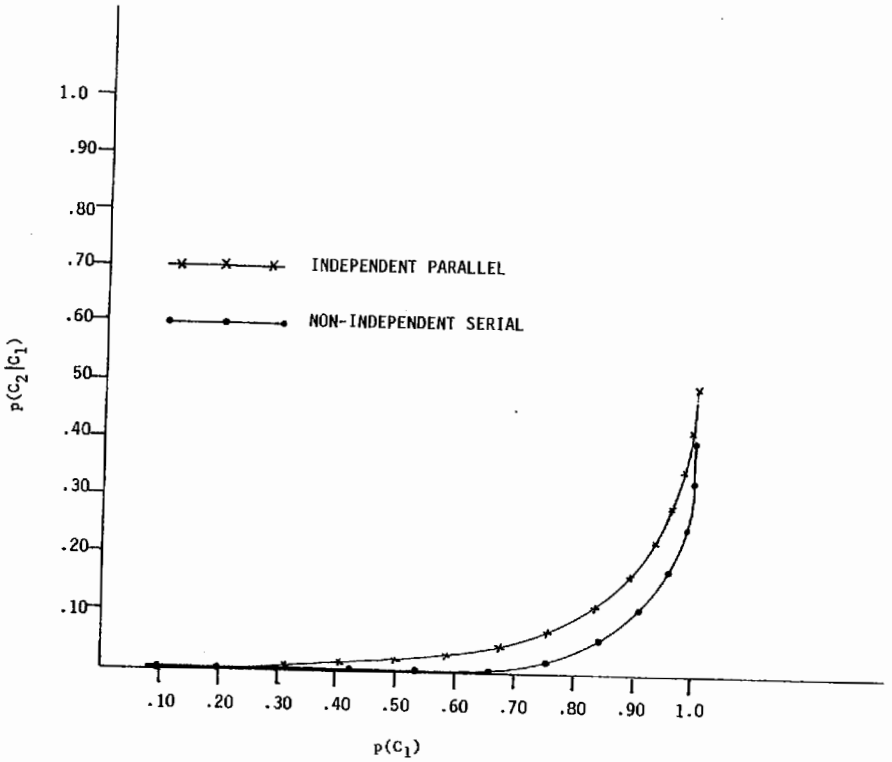


Fig. 11.4. The conditional probability correct in position 2 as a function of probability correct on position 1 for a fixed-order serial and an independent parallel model (replotted from Fig. 11.3).

completed, position 1 must have been finished. Therefore, $P(C_2 \mid C_1) = P(C_2)/P(C_1)$. The result is shown in Fig. 11.4. The parallel and serial predictions are qualitatively similar and the maximum difference between the two is .08, whereas it is about .10 in the unconditional $P(C_2)$ curves of Fig. 11.3. Thus, the parallel–serial difference is lessened in this method of plotting. Further, if the v_2 parallel rate has specifically been selected so that the serial position 2 function would be more closely mimicked (e.g., by picking $v_2 < v/2$), the conditional serial curve would actually be greater than the parallel at some points.

Emphasizing the importance of the independence dimension, we find that a fixed-order serial model with negative correlations can, in contrast to the above outcome, even lead to an *accentuated* difference.

Degradation by masking in serial and parallel systems

In the last comment before we introduce the details of the experimental study, it is of interest to examine briefly the degradation expected with shielded dis-

play positions in a stochastic serial model. (A deterministic fixed-order serial model will, as noted, predict virtually no degradation.)

We are unfortunately not in a position at present to make any completely general statement. However, we can illustrate the situation within a common set of models. For expository purposes, we will confine our attention to a simplified right shield (RS) condition in which a shield (opaque but blank mask) is placed over a rightmost letter in a two-letter display. The shield is assumed to be removed after a duration that, in a left-to-right serial system, should have allowed the first letter to be completed most of the time. Thus, the processor should be ready to work on the second letter when it comes on. Conversely, it is supposed that in the absence of the shield (no shield = NS), such a serial processor should have little or no more time to work on the second letter than it would in the RS condition.

Suppose that in NS the letters are displayed for t msec and that in RS the left letter is also shown for t msec. The right letter, however, comes on for a duration of t_R msec after a period of $t-t_R$ msec from the start of the display proper (which begins with the presentation of the left letter).

Now consider the action of an exponential serial model; in particular, we will concentrate on the right position. In this model the right letter can be correct only if both letters are completed. In the NS condition, this probability is

$$P(C_2 \mid \text{NS}) = \int_0^t \int_0^{t-t_2} u \exp(-ut_1) u \exp(-ut_2)\, dt_1\, dt_2$$
$$= 1 - \exp(-ut) - ut \exp(-ut)$$

where t_i is the actual processing time or the intercompletion time associated with position $i = 1$ or 2. Note that the above result is equivalent to the Poisson probability that at least two items are completed [$1 - P(0$ are completed$) - P(1$ is completed)].

The probability that position 2 is finished in the RS condition can be calculated as a function of the probability that the first is done before the second even comes on (i.e., in the interval $t-t_R$) times the probability that the second is completed in its alloted time (t_R). To this quantity is added the integral of the density that the first is done at some time t_1 after $t-t_R$, but before t times the probability that the second is finished in the remaining interval $(t-t_1)$. This amounts to

$$P(C_2 \mid \text{RS}) = \{1 - \exp[-u(t-t_R)]\}[1 - \exp(-ut_R)]$$
$$+ \int_{t-t_R}^{t} u \exp(-ut_1)\{1 - \exp[-u(t-t_1)]\}\, dt_1$$
$$= \{1 - \exp[-u(t-t_R)]\}[1 - \exp(-ut_R)]$$
$$+ \{\exp[-u(t-t_R)] - \exp(-ut) - ut_R \exp(-ut)\}$$
$$= 1 - \exp(-ut_R) - ut_R \exp(-ut)$$

The degradative difference $P(C_2 \mid \text{NS}) - P(C_2 \mid \text{RS})$ will now be calculated for later comparison with the analogous parallel difference:

$$\Delta_s = P(C_2 \mid \text{NS}) - P(C_2 \mid \text{RS}) = \exp(-ut_R) - \exp(-ut) - u(t-t_R)\exp(-ut)$$

The independent parallel equations are simplicity itself. The probability of being correct on the second letter in condition NS is just $P(C_2 \mid \text{NS}) = 1 - \exp(-vt)$ and in condition RS is $P(C_2 \mid \text{RS}) = 1 - \exp(-vt_R)$, because it does not matter, due to independence (of individual total completion times; see Chapter 4), what happens on position 1.

Therefore, the pertinent degradative difference is

$$\Delta_p = \exp(-vt_R) - \exp(-vt)$$

Proposition 11.1: Suppose that $t_R = t/2$ in the simplified paradigms; that is, the shield goes off halfway through the total duration of presentations. Then the degradation caused by the RS condition in independent exponential parallel processing is always greater than that caused in fixed-order exponential serial processing.

Proof: In order to attain a realistic comparison of the relative degradation predictions we need to equate $P(C_2 \mid \text{NS})$ in the two models [note that the parallel $P(C_1 \mid \text{NS})$ could also be set equal to the serial $P(C_1 \mid \text{NS})$ by selecting a distinct rate parameter $v_1 \neq v_2$]. We thus set

$$P(C_2 \mid \text{NS; serial}) = 1 - e^{-ut} - ute^{-ut} = 1 - e^{-vt} = P(C_2 \mid \text{NS; parallel})$$

The parallel parameter v can now be solved for u in terms of the serial prediction that will guarantee the above identity for any serial rate u and duration t.

Hence, we discover after a little simplification that

$$v^* = u - \frac{1}{t}\ln(1+ut) \geqslant 0$$

It should not be thought that it is being proposed that v^* is a function of time in the parallel model. Rather, this equation simply demonstrates that the parallel and serial predictions can be equated for any given duration t.

We are now in a position to compare Δ_s and Δ_p. First note that while Δ_s is as above, we will now write $\Delta_p = \exp(-v^*t_R) - \exp(-v^*t)$. Now let $t_R = t/2$, as assumed in the statement of the proposition. Then the degradation comparison can be expressed as

$$\Delta_p - \Delta_s = \exp\left\{-\left[u - \frac{1}{t}\ln(1+ut)\right]\frac{t}{2}\right\} - \exp\left\{-\left[u - \frac{1}{t}\ln(1+ut)\right]t\right\}$$
$$- \left[e^{-ut/2} - e^{-ut} - ue^{-ut}\left(t - \frac{t}{2}\right)\right]$$
$$= e^{-ut/2}[(1+ut)^{1/2} - 1] - \frac{ut}{2}e^{-ut}$$

This difference will be positive if and only if

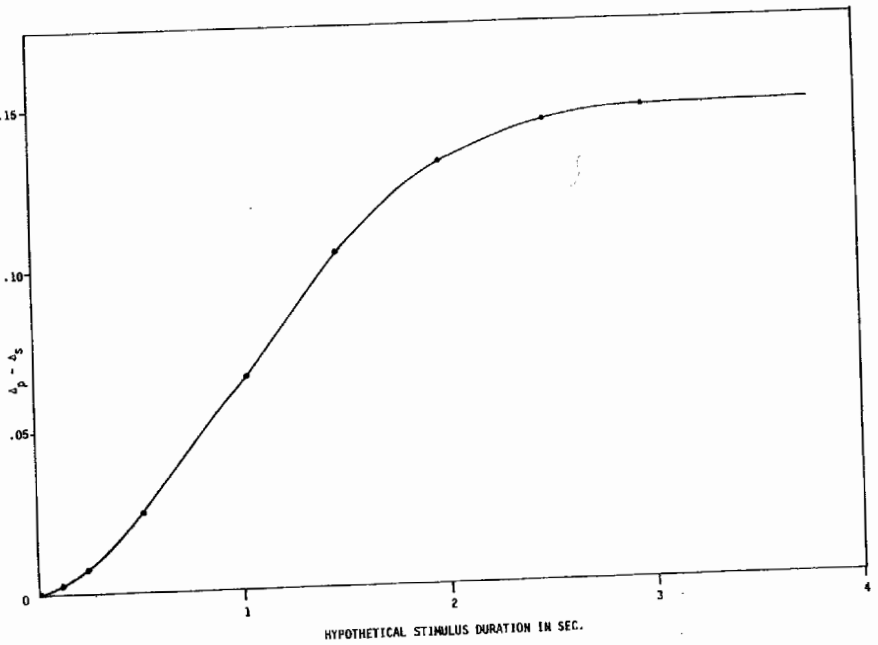


Fig. 11.5. The predicted difference in performance degradation between a serial and parallel model (Δp = predicted parallel degradation; Δs = predicted serial degradation).

$$e^{-ut/2}[(1+ut)^{1/2}-1] > \frac{ut}{2}e^{-ut}$$

Now square both sides and simplify some more to arrive at

$$1 > e^{-ut/2} + \frac{ut}{4}e^{-ut}$$

The biggest the right-hand side can ever be is 1, at $ut=0$, and it decreases thereafter. It can therefore be concluded that $\Delta_p > \Delta_s$ always; that is, the parallel system will always show more degradation than the serial system. □

Suppose it takes on the average 15 msec to process a letter; that is, in exponential serial terms, $1/u = 15$ msec or $u = .067$ letters/msec. In fact, this is very close to the observer average estimated from the data discussed in the following section. (The range was 11.5–23.5 msec.) Then Fig. 11.5 shows $\Delta_p - \Delta_s$ as t runs from 0 to 40 msec. Observe that with a stimulus duration of twice the average processing time ($t = 30$ msec), the curve is close to its maximum. At $t = 30$ msec, Δ_p is more than twice Δ_s. If Δ_p were actual data with a sample size, say, of 75, a t test of the serial prediction minus the (parallel) data, $\Delta_s - \Delta_p$, would yield significance at about the $\alpha = .05$ level.

It is difficult to generalize in detail to more complex experiments with larger sets of letters. However, if the right shield were to cover approximately $n/2$ letters, the peak difference $\Delta_p - \Delta_s$ would be approximately achieved by presenting the stimuli (not counting the shield) for about n times the estimated $1/u$. This duration turns out to be approximately that employed in Phase 2 below.

This example seems instructive as well as propitious, but further mathematical work needs to be done. The left shield effects should be predictable by symmetry, but n needs to be varied and placement of the shield studied with respect to maximizing $\Delta_p - \Delta_s$.

A whole report experiment

The stimuli consisted of 2,000 sets of five letters each.[1] Only consonants were used. These five-letter sets were determined by programming a computer to randomly draw five letters at a time, without replacement, from the set of 20 consonants. Then, the five letters were "replaced" into the original set of 20 and the process begun again. This was done 2,000 times, producing 2,000 sets of five letters each. The stimuli were typed, each set of five letters being placed on 2″ × 2″ white cards. The arrays as viewed yielded letters subtending heights of 23 min, with the entire five letters being about 2.5 deg visual arc.

A Gerbands model T-2B two-field tachistoscope was modified to yield three independent fields by means of solenoid switching of the first (which then became the third) stimulus field. The modifiable field could thus be either a white field with a .2-cm-diameter fixation dot visible through a 1.8-cm-high by 6-cm-wide aperture in the center (field 1), or the same white field with a scrambled letter "visual noise" pattern visible through and filling the aperture (field 3). The stimulus field (field 2) was a white field with a 1.8-cm-high by 3.6-cm-wide aperture in the center. Through this aperture could be seen the five-letter stimulus pattern on any one trial. Another piece of apparatus allowed the first two letters on the left or the last two letters on the right of the stimulus to be covered for various portions of the stimulus presentation interval (used only during Phase 2). It consisted of a high-speed shutter mechanism that operated either a left shield (LS) or right shield (RS) or was

[1] The original experiment reported here was run at the University of Hawaii in 1967 and reported by Townsend and Fial (1968). Part of the experiment (Phase 2) involves the idea of manipulating the presence or absence of selected letters in the display. In that way it is similar to the withdrawal of a selected letter midway during its exposure that Sperling (1970) used about the same time. Our theoretical conceptions concerning the present type of phenomena have continued to mature over the years, and resulted in the present chapter and a significantly more terse rendition of the study in *Acta Psychologica* (1981). The experiment also bears similarity to the forced serial processing technique of Travers (1973). Performance decrements were found with word but not with random letter strings. The latter result contrasts with the substantial shield effects found in Phase 2 of the present study.

inoperative on any one trial. The portion of the shield in the observer's view was made of the same material that the stimulus was printed on and was white and blank. The brightness of the fields was 25.35 cd/m^2. The ambient illumination was 19.53 cd/m^2 at the observer's position. The opaque 2″ × 2″ cards were presented via an automatic changing mechanism in randomly arranged batches of 100. A button was made available to the observer for self-presentation of the stimulus, which occurred after a 1-sec delay.

The total experiment required 10 days for each observer, Phase 1 occupying 8 days and Phase 2 the other 2.

Phase 1

When the observer pushed the initiation button the white prestimulus field (field 1) with the fixation point was replaced after 1 sec by the stimulus field (field 2) containing five letters slightly above the locus of the fixation point. At the termination of the display duration, the stimulus field was replaced by the visual noise field (field 3) for 500 msec, after which the observer attempted to verbally report the letters in the display from left to right. He or she was then told the letters that had appeared in the display. After a day of instruction and practice, Days 2 through 8 contained 200 trials each (after a set of warmup trials), 100 trials at each of two stimulus durations. Day 2 tested durations of 250 and 200 msec, whereas Day 3 tested 170 and 150 msec and Day 4 tested 120 and 100 msec. Days 5 through 8 descended from 90 msec in 10-msec decrements to 10 msec during the second half of Day 8.

Phase 2

Day 9 was taken up with practicing the observer in the shield conditions, which consisted of LS (left shield), RS (right shield), or NS (no shield), and brief testing at the 15 stimulus durations employed in Phase 1. Plots were made of each observer's average performance on each of the 15 stimulus durations used in Phase 1, and estimates from these curves of the minimum amount of time required for two and four letters to be correctly reported were calculated separately for each observer. This information was compared with the Day 9 test results to determine the final values. The four-letter duration (referred to as $t4$ msec) was employed as the basic stimulus presentation interval for Day 10, and the two-letter duration ($t2$) was employed as the shield interval. It should be noted that $t4$ was not constrained to be twice $t2$.

In the LS condition, all five letters were presented, then the first two on the left were covered up after the first $t2$ msec had elapsed, but the rightmost three letters stayed uncovered the whole $t4$ msec of the presentation interval. In the RS condition, the rightmost two letters of the stimulus were covered for the first part of the $t4$ msec of the stimulus duration, then uncovered for the last $t2$ msec of the stimulus period. The shield itself was blank to preclude additional lateral interference effects on the still-visible positions 3, 4, and 5

in LS. This resulted in a slight advantage in poststimulus conditions for positions 1 and 2 in LS as opposed to 4 and 5 in RS, which ended with the global poststimulus noise mask.

On Day 10, in addition to warmup trials with the various conditions, 75 trials were given at each of the three shield conditions (LS, RS, and NS), with their order being randomly determined for each observer.

Eight naive undergraduate students were run as observers and were paid for their participation in the experiment. All had 20/20 acuity after correction, and those wearing corrective lenses wore them throughout the study.

Analysis, results, and discussion

First, the results and preliminary discussion of the serial position curves, the independence analyses, and the shield conditions are separately presented. This is followed by a general discussion that attempts to develop a plausible tentative model of whole report processing.

Phase 1: The serial position curves

Figure 11.6a shows the serial position curves obtained in Phase 1 averaged over all eight observers, and Figs. 11.6b through 11.6i indicate the curves for the individual observers. As is suggested by these figures, the results were quite consistent across observers and replicated Sperling's (1967) findings. In addition to fanning out from a common origin, the probability-correct curves are not changing appreciably at durations longer than 100 msec. In fact, the curves of most observers reach asymptote, but not perfect performance, by 40 or 50 msec. At about that duration, a limit appears to be reached, presumably in identification, translation, or immediate memory capacity or all three, and this limitation is not compensated for even by an extra 200 msec display time. As it turned out, the basic form of these curves was affected by whether we required a correct answer to include the correct positioning of the letter in the response sequence. All the plots and analyses concerned hereafter were performed for identification accuracy only, without regard to specific position.

A priori, the descending stimulus duration sequence might have interacted with session number and elevated slightly the latter accuracy levels (through practice). However, the data taken in session 9 (Day 9) on the 14 durations yielded the same basic functions, with little elevation of probability correct. In addition, the rapid approach to asymptote also appears in Sperling's (1967) curves.

Obvious in every observer's curves was a left-to-right ordering of the curves with a reversal of serial positions 4 and 5. Thus, potentially, processing might have been left-to-right and serial, or parallel with some positions being processed more accurately (e.g., more quickly) than others. The reversing of positions 4 and 5 may be due to an interaction between attention and lateral

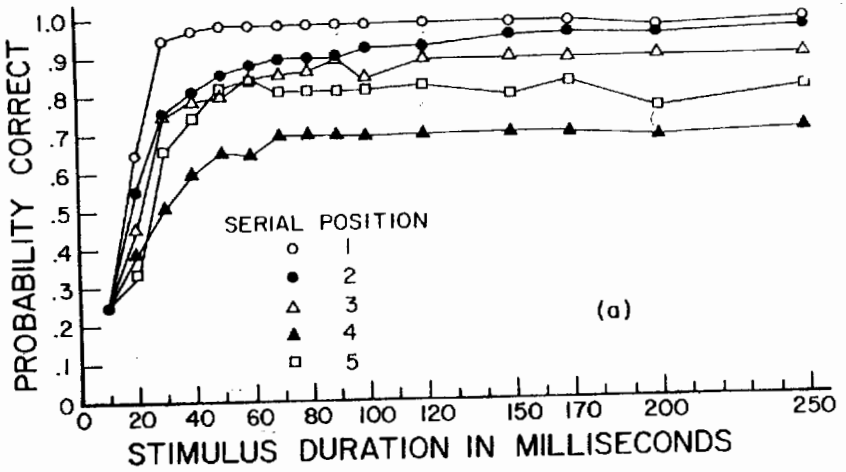
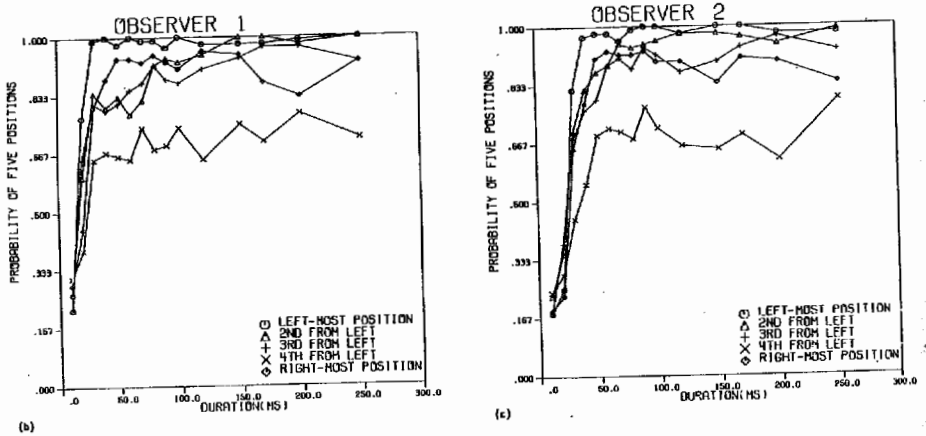
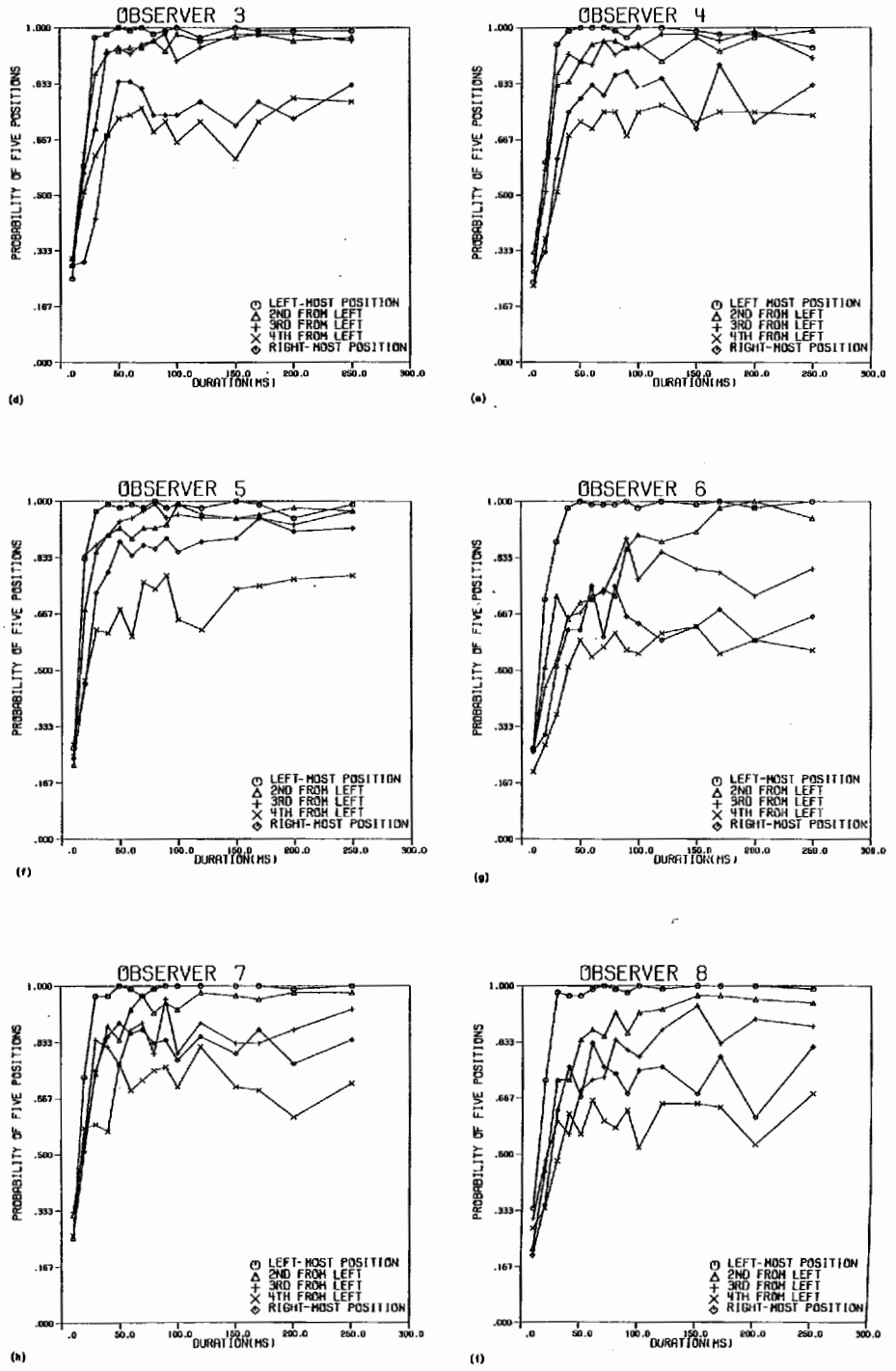


Fig. 11.6(a). Group averaged probability-correct curves for serial positions 1–5.



Figs. 11.6(b)–(i). Individual observer probability-correct curves for positions 1–5.



interference or other retinal acuity factors (Townsend, Taylor, & Brown 1971; Estes, Allmeyer, & Reder 1976) or to attentional factors alone. If the observers held their gaze at the centered fixation point, the reversal cannot have been due only to retinal factors.

Although our earlier arguments demonstrated that serial as well as parallel models could exhibit "fan" behavior, we have at this point at least a firm replication of Sperling's findings, with a very salient aspect being the clear upper bound on performance that appears to be reached relatively early.

Although we will not be able to draw absolute conclusions regarding parallel vs. serial processing in the present study, we will see that overall we can further delimit the type of model that can satisfy the data. For instance, independence of processing is more intuitively associated with parallel than serial processing, although some serial models can produce this characteristic. We therefore next analyze the data pertinent to this issue, which is also important in its own right.

Phase 1: Independence vs. dependence of processing

In order to gain an idea of the interletter dependencies in the various positions, predictions were obtained from several diverse models and compared with the data.

As noted above, serial models based on the Poisson distribution produce positive dependencies. Such models are compatible with an overall theory positing time-limited identification and/or translation to verbal-acoustic immediate memory with these constraints being sufficiently severe that immediate memory is not taxed. On the other hand, if the primary limitation were in immediate memory (afs) and this memory were of fixed capacity, then a negative correlation would result. The reason is that if it is known that a given letter has made it into immediate memory, then the chances are lessened that space will be available for another letter. It would further seem that other theories positing output interference as the limiting factor in whole report performance might also predict negative correlations, but apparently none has been sufficiently developed to make precise predictions on this issue. An independence result seems intuitively most compatible with a set of independent channels in vfs.

Two positive dependence models, a negative dependence model, and an independence model were employed to make dependence predictions. Each model contained guessing structure appropriate to its particular characterization.

The processing (as opposed to guessing) part of the positive dependency models was based on a fixed-order (left-to-right) serial Poisson model and a random-order serial Poisson model, respectively. The negative dependence model is mathematically equivalent to the Estes and Taylor (1964) fixed-sample-size model and will henceforth be so designated. In contrast, the independence model assumed that each letter was processed independently of the others.

The mathematical form of these models can be indicated by the form of their probability-correct $[P(C_i)]$ prediction for a given display position i.

The analysis of independence below involves averaging across display positions, and some of the model predictions can ignore these. However, as we will see, the left-to-right Poisson model has to take position into account by its very nature. We will also derive serial position predictions of the other models. It should be emphasized that the $P(C_i)$ equations do not themselves exhibit the interpositional dependencies or lack thereof. The actual dependency predictions will be developed later.

For all models, n=display size and N=stimulus population size. In the present experiment, $n=5$ and $N=20$.

The probability correct on position i in the case of the strict left-to-right serial Poisson model is

$$P(C_i)=\sum_{j=0}^{i-1} P(j,t)\,\frac{n-j}{N-j}+\sum_{k=i}^{\infty} P(k,t)$$

where

$$P(j,t)=\frac{(ut)^j e^{-ut}}{j!}$$

is the Poisson probability that j letters have been completed by time t, and $(n-j)/(N-j)$ is the probability of correctly guessing the letter in position i when j letters have been completed. Thus, the first sum gives the contribution of trials where the display was terminated before reaching stimulus location i. For instance, let us look at $P(C_3)$, the probability correct in stimulus position 3. Suppose that during the display time t, the first two ($j=2$) letters were completed in the left-to-right processing path. This occurs with probability

$$P(2,t)=\frac{(ut)^2 e^{-ut}}{2!}$$

Because neither of these two letters will be in the guessing pool that exists for the unidentified remaining three letters, the average probability of correctly guessing the letter that was in position 3 is simply

$$\frac{5-2}{20-2}=\frac{3}{18}=.167$$

The product

$$\frac{(ut)^2 e^{-ut}}{2!}\cdot(.167)$$

will therefore be the second term in the first sum above. The second sum picks up those terms wherein position 3 is completed so that the $P(j,t)$ $(j\geq 3)$ is simply multiplied by 1. Adding together both sums yields the overall average probability correct. Note that this model assumes the completion time distribution for each letter is exponential with processing rate u. Serial position effects are represented by $P(C_i)>P(C_j)$ if $i<j$.

The random-order serial model predicts that

$$P(C_i)=P(C)=\frac{1}{n}\left\{\sum_{m=1}^{n}\left[\sum_{j=0}^{m-1}P(j,t)\,\frac{n-j}{N-j}+\sum_{k=m}^{\infty}P(k,t)\right]\right\}$$

that is, $P(C)$ is an average of the possible values of probability correct across the different processing positions. Because the processing path through the various display positions is presumed to be entirely random, it follows that any position (i) will appear in any one of the possible n processing locations with probability $1/n$. The index m in the above equation represents this (mth) processing location. We can thereby represent the probability correct on display position i as

$$P(C_i)=\sum_{m=1}^{n}P(\text{position } i \text{ is processed } m\text{th})\cdot P(\text{correct given position } i \text{ is processed } m\text{th})$$

where $P(\text{position } i \text{ is processed } m\text{th})=1/n$.

Given the processing position m, the above conditional probability is then clearly exactly the same as $P(C_m)$ in the previous formula for the strict left-to-right processing path; that is,

$$P(C_m)=\sum_{j=0}^{m-1}P(j,t)\,\frac{n-j}{N-j}+\sum_{k=m}^{\infty}P(k,t)$$

$P(C_i)$ then results as the weighted sum of these $P(C_m)$ terms. This model predicts a complete absence of serial position effects.

Consider next the independence model. The general form of the probability correct formula is

$$P(C_i)=P(C)=p(t)+[1-p(t)]\,\frac{n}{N},\quad 0\leqslant p(t)\leqslant 1$$

The term $p(t)$ is the probability that the letter in a position is correctly processed during the display time t. With probability $1-p(t)$ the position is not completed and the observer must guess the entry. Because we do not know what happens in the other positions, the overall average probability correct by guessing in that event is just n/N. Note that $p(t)$ is fixed for a given t, but we expect it to increase as t increases. Serial position effects can be implemented in this model by simply letting $p_i(t)\neq p_j(t)$ for $i\neq j$.

The fixed-sample-size model makes the prediction

$$P(C)=\frac{\binom{1}{1}\binom{n-1}{X_t-1}}{\binom{n}{X_t}}+\frac{\binom{1}{0}\binom{n-1}{X_t}}{\binom{n}{X_t}}\,\frac{\binom{1}{1}\binom{N-1-X_t}{n-1-X_t}}{\binom{N-X_t}{n-X_t}},\quad 0\leqslant X_t\leqslant n$$

The appropriate probability distribution for this model is, as shown, the

number of items of a particular variety when a fixed sample is drawn from a set of items. When the stimulus duration is t, the fixed sample size is X_t (a positive integer), which of course must be of size $0\leqslant X_t\leqslant n$. We are interested first in the number of ways in which the particular item of interest, namely the letter in position i, can be drawn in this selection of X_t letters out of the n. This is given by

$$\binom{1}{1}\binom{n-1}{X_t-1}$$

and the probability the ith letter is sampled is just

$$\frac{\binom{1}{1}\binom{n-1}{X_t-1}}{\binom{n}{X_t}}$$

that is, the above term divided by the total number of ways of sampling X_t from n items. This term is then the probability of being correct on a position by actual identification.

However, if the letter in that location is not contained in the sample of X_t letters, which happens with probability

$$\frac{\binom{1}{0}\binom{n-1}{X_t}}{\binom{n}{X_t}}$$

then one may still be correct by guessing. Unlike the situation in the independence model, we know that exactly X_t of the n displayed letters have been perceived, so the guessing probability must take account of this fact. The appropriate latter probability is the hypergeometric probability that a certain one (namely the one in position i) of the remaining $n-X_t$ letters is sampled from the total of $N-X_t$ possibilities. Finally, although X_t is fixed for a given t, we expect it to increase as t becomes larger. As written, this model cannot reflect serial positions, although a much more complicated fixed sample size model could by assuming that some positions can be sampled more easily than others.

Analysis of the statistics $P(C_i)-P(C_i\,|\,C_j)$ and $P(C\,|\,k)$

A statistic that exhibits the degree of interdependence present is

$$\text{ave}[P(C_i)-P(C_i\,|\,C_j)]$$

where $i\neq j$ and the average is taken over i and j. By plotting this quantity as a function of ave$[P(C_i)]$, it is possible to compare the data with predictions of the four models in a way that does not depend on specific parameter values

As a labor saver for other investigators who might have occasion to carry out independence analyses analogous to those employed here, the formulas will be given for the four models.

In all of the models we already know $P(C_i)$, and because

$$P(C_i \mid C_j) = \frac{P(C_i \cap C_j)}{P(C_j)}$$

the only probability remaining to be derived is $P(C_i \cap C_j)$. To produce an average overall statistic, we compute the simple average

$$\frac{1}{5 \times 4} \sum_{i=1}^{5} \sum_{\substack{j=1 \\ \neq i}}^{5} [P(C_i) - P(C_i \mid C_j)]$$

and in the case of the data the above quantity was also averaged over all observers.

Consider first the fixed-order Poisson model. Observe that if $i<j$, then one may be correct on both positions by completing neither and guessing both, by completing i and guessing j, or by completing both. Thus, we write, without loss of generality, for $i<j$,

$$P(C_i \cap C_j) = P(0,t) \frac{n}{N} \cdot \frac{n-1}{N-1} + P(1,t) \frac{n-1}{N-1} \cdot \frac{n-2}{N-2} + \cdots + P(i,t) \frac{n-i}{N-i} + \cdots + \sum_{k=j}^{\infty} P(k,t)$$

where of course the term with $P(i,t)$ is the first where the observer only has to guess one of the two positions. For any given i and j the quantity $P(C_i) - P(C_i \mid C_j)$ easily follows from the above computation.

The random order Poisson model's predictions are very similar except that it is necessary to average over the various potential *processing* positions of the stimulus locations i and j; thus, $P(C_i \cap C_j)$ is an average of all possible fixed-order terms $P(C_k \cap C_l)$, for any i,j:

$$P(C_i \cap C_j) = \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{\substack{l=1 \\ \neq k}}^{n} P(C_k \cap C_l)$$

where $P(C_k \cap C_l)$ can be calculated from the fixed-order Poisson formula. Similarly,

$$P(C_i) = P(C) = \frac{1}{n} \sum_{k=1}^{n} P(C_k)$$

Next, consider the independence model. Again we are concerned only with the two positions i and j, either of which can be reported accurately either by guessing or correct processing:

$$P(C_i \cap C_j) = p^2(t) + 2p(t)[1-p(t)] \frac{n-1}{N-1} + [1-p(t)]^2 \frac{n}{N} \cdot \frac{n-1}{N-1}$$

Finally, the fixed-sample-size model predicts that $P(C_i \cap C_j)$ is produced by the three possibilities: (1) getting both in the fixed sample X_t (which has probability greater than 0 if $X_t \geq 2$); (2) getting only one in X_t; or (3) getting neither in X_t. When one or both *are not* included in X_t, it may still be guessed correctly.

Hence,

$$P(C_i \cap C_j) = \frac{\binom{2}{2}\binom{n-2}{X_t-2}}{\binom{n}{X_t}} + \frac{\binom{2}{1}\binom{n-2}{X_t-1}}{\binom{n}{X_t}} \cdot \frac{\binom{1}{1}\binom{N-1-X_t}{n-1-X_t}}{\binom{N-X_t}{n-X_t}} + \frac{\binom{2}{0}\binom{n-2}{X_t}}{\binom{n}{X_t}} + \frac{\binom{2}{2}\binom{N-2-X_t}{n-2-X_t}}{\binom{N-X_t}{n-X_t}} \quad \text{if } X_t \geq 2$$

If $X_t=0$, then

$$P(C_i \cap C_j) = \frac{n}{N} \cdot \frac{n-1}{N-1}$$

that is, guessing provides the only possibility for being correct, and if $X_t=1$, then

$$P(C_i \cap C_j) = \frac{\binom{1}{1}}{\binom{n}{1}} \cdot \frac{\binom{1}{1}\binom{N-2}{n-2}}{\binom{N-1}{n-1}} + \frac{\binom{1}{0}}{\binom{n}{1}} \cdot \frac{\binom{2}{2}\binom{N-3}{n-3}}{\binom{N-2}{n-2}}$$

where the first term on the right supposes either position i or j is the sampled one and the second presumes that both have to be guessed. The overall statistic can now be put together with these plus the $P(C_i)$ terms. Now to the data.

Figure 11.7 shows these predicted functions, along with the data points of the eight individual observers presented as a scatter plot. The display time, of course, is an implicit parameter here since the different durations caused ave$[P(C_i)]$ to vary. The data points were generated by blocking the display times into six intervals, 10 msec, 20 msec, 30 msec, 40–50 msec, 60–100 msec, and 120–250 msec, and then averaging $P(C_i)$ and $P(C_i) - P(C_i \mid C_j)$, as noted, within these intervals and over stimulus positions i and $j=1$ to 5. The fixed-order and random-order Poisson models generated sufficiently similar predictions that only the fixed-order curve is graphed.

Figure 11.7 exhibits the fixed-sample-size model prediction of a large negative correlation, the independence model prediction of a small negative correlation due to guessing, and the Poisson model prediction of a sizable positive correlation, at least for $P(C_i) > .30$. The data are much closer to the independence predictions than either of the other two model predictions.

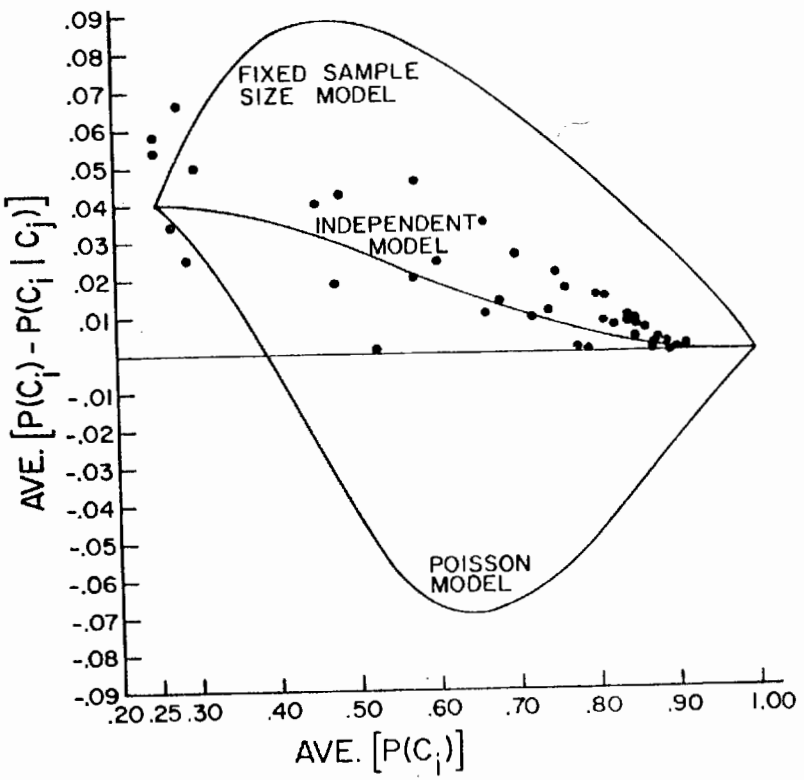


Fig. 11.7. An interletter dependency statistic $P(C_i)-P(C_i|C_j)$, as a function of $P(C_i)$, averaged over serial positions (i,j) and observers.

Individual observer's curves appeared to agree with the overall scatter plot results of Fig. 11.7, and none showed a consistent tendency to deviate toward the positive or negative correlation predictions. For instance, when $P(C_i) \geq .80$, no observer possessed absolute curve values above .02, whereas the fixed-sample-size model still predicted relatively large positive deflections and the Poisson model relatively large negative deflections from zero. These tendencies also held up in the many individual observer plots of $P(C_i)-P(C_i|C_j)$ for each $1 \leq i \leq 5$ across the serial position $j \neq i$ within each single stimulus duration. That is, $P(C_i)-P(C_i|C_j)$ tended to go from a positive value toward 0 as the duration [and therefore $P(C_i)$] increased. A representative example is shown in Fig. 11.8, the data of observer 5 in the case of stimulus position $i=2$. Here $P(C_2)-P(C_2|C_j)$ is on the ordinate and j is on the abscissa.

Another statistic for testing independence is simply $P(C|k)$, the probability correct given exactly k of the other positions were correct, averaged again over the five serial positions and plotted against k. The predictions of

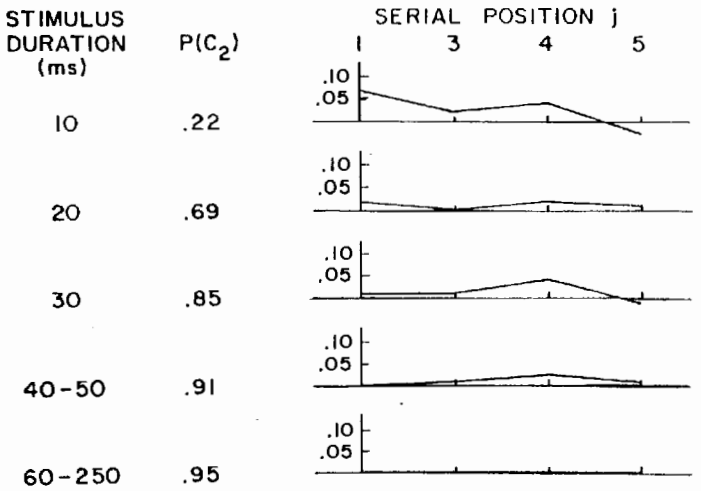


Fig. 11.8. $P(C_2)-P(C_2|C_j)$ plotted as a function of serial position of other letters $(j=1,3,4,5)$ as stimulus duration is varied for an individual observer.

the various models may be worked out in analogous fashion to the earlier conditional statistic, except that now one simply counts the number of other letters that are correct, irrespective of their stimulus positions. The results of this analysis are shown in the panels of Fig. 11.9a–e. Data and predictions are omitted where few trials contributed to the point. The different panels show this statistic for varying exposure duration, grouped to contain equally accurate data. Theoretical predictions were generated by selecting those parameter values yielding the same marginal probability correct as for the particular accuracy level. Thus, each model had one parameter estimated for each graph in Fig. 11.9 or five in all. Note that here, the random Poisson and the fixed-path Poisson models make quite distinct predictions. Clearly, the data are again much closer to the independence model than to any of the others.

Phase 2

The shield results are shown in Table 11.1. The theoretical entries (THE) will be discussed later. Noting that the maximum standard error for any proportion in Table 11.1 is $(.25)^{1/2}/(75)^{1/2}=.06$ (from $[p(1-p)]^{1/2}/N^{1/2}$, it can be seen that the shield caused sizable effects, typically deterioration in the shielded positions. For instance, performing a sign test on the 32 differences taken from $NS(i)-LS(i)$, $i=$position 1 or 2 and $NS(i)-RS(i)$, $i=$position 4 or 5 for the individual observers, finds significance exceeding the .001 level.

It was expected that the LS effects would not be as strong as the RS effects since the former type of trial was terminated by the blank shield, but the RS

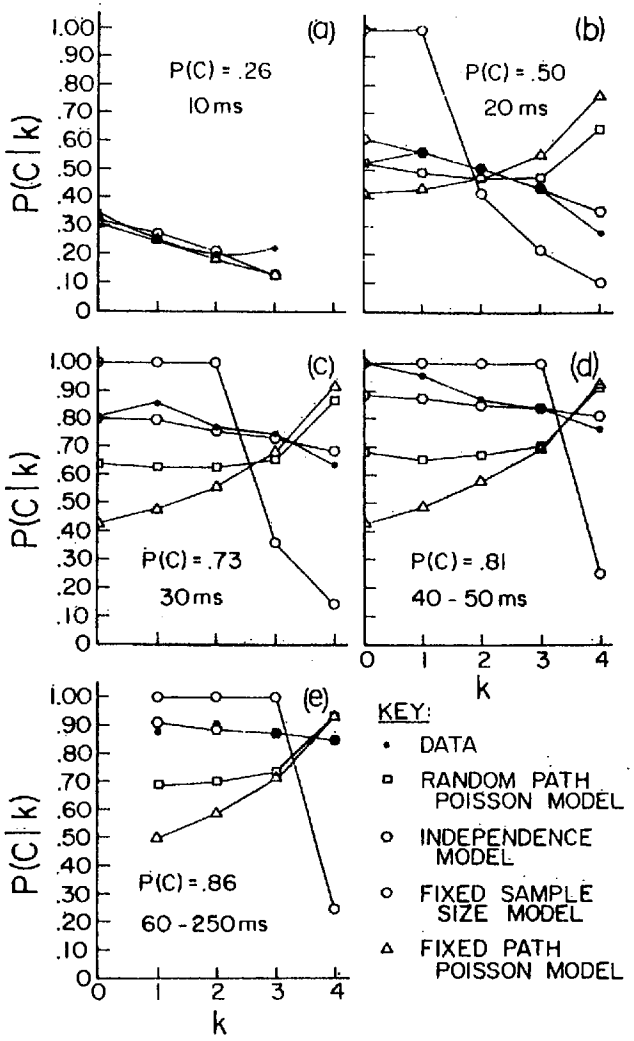


Fig. 11.9. The interletter dependency statistic $P(C|k)$ as a function of k, averaged over serial positions and observers, for different marginal probability-correct values.

trials were terminated by the usual visual-noise poststimulus mask. This did occur; indeed in most cases, serial position 1 showed little information loss.

The other second-order effect of note is the appearance of some individual differences in reaction to the shields. Some observers seem to have let accuracy on the right positions suffer somewhat in LS in order to perform well on the first two positions, especially the first. Observer 4 actually improved performance on the first three positions (position 3 was not shielded in either

Table 11.1. *Results of the shield conditions*

		Serial position 1			Serial position 2			Serial position 3			Serial position 4			Serial position 5		
S		NS	RS	LS	NS	RS	LS	NS	RS	LS	NS	RS	LS	NS	RS	LS
1	OBS[a]	.99	.99	.93	.71	.85	.67	.80	.92	.68	.64	.52	.63	.76	.35	.68
	THE	1.00	1.00	.93	.81	.81	.46	.79	.79	.79	.63	.39	.63	.79	.45	.79
2	OBS	.99	1.00	.95	.84	.84	.77	.84	.85	.72	.73	.31	.60	.73	.21	.85
	THE	.97	.97	.76	.79	.79	.63	.75	.75	.75	.58	.49	.58	.73	.59	.73
3	OBS	.97	.98	.87	.92	.92	.72	.96	.93	.89	.79	.71	.77	.80	.69	.84
	THE	1.00	1.00	.87	.92	.92	.54	.92	.92	.92	.72	.45	.72	.74	.46	.74
4	OBS	.91	.99	1.00	.79	.89	.84	.88	.91	.92	.80	.68	.69	.85	.61	.81
	THE	1.00	1.00	1.00	.91	.91	.51	.90	.90	.90	.70	.43	.70	.78	.46	.78
5	OBS	1.00	1.00	.95	.83	.80	.76	.95	.99	.93	.79	.48	.67	.81	.55	.79
	THE	1.00	1.00	.95	.92	.92	.49	.94	.94	.94	.69	.41	.69	.84	.46	.84
6	OBS	.99	.97	1.00	.80	.79	.63	.89	.71	.84	.61	.76	.63	.76	.80	.75
	THE	1.00	1.00	.85	.84	.84	.54	.77	.77	.77	.59	.42	.59	.68	.46	.68
7	OBS	1.00	.99	.97	.88	.88	.81	.89	.83	.88	.76	.71	.72	.93	.85	.84
	THE	1.00	1.00	1.00	.90	.90	.88	.84	.84	.84	.70	.68	.70	.82	.81	.82
8	OBS	.99	.96	.96	.84	.81	.72	.55	.77	.67	.64	.56	.63	.76	.82	.83
	THE	1.00	1.00	.75	.82	.82	.48	.72	.72	.72	.75	.38	.75	.68	.43	.68
AVE	OBS	.98	.99	.96	.83	.85	.74	.85	.86	.82	.72	.59	.67	.80	.57	.80
	THE	1.00	1.00	.89	.86	.86	.57	.83	.83	.83	.67	.46	.67	.76	.52	.76

[a] OBS = observed; THE = theoretical.

condition) in LS with some lowered probability correct in positions 4 and 5. Some observers seemed to compensate for the shield even in RS with more attention on one or both of the shielded positions, the most noteworthy result being perhaps the substantial increase in probability correct on position 4 by observer 6 in RS; it may be observed that there was a simultaneous loss at this observer's unshielded position 3. Thus, it seems that some observers can alter their interposition distribution of attention between blocks when the viewing environment is changed and may do so with some between-individual variation.

One additional point is important here. It is that the deterioration in accuracy occurred even though the design used here employed only a single shield per trial rather than two, one on each end. In that situation, the right shield would be on when the display is revealed, then shortly after it was removed the left shield would be brought in until the entire display was terminated with the visual noise. The latter design would be expected to cause even more degradation of performance in parallel processing.

The shield effects seem most compatible with a parallel system that can adjust attention on a between-block basis. There appears to be no single observer in Table 11.1 who gives evidence entirely compatible with a simple left-to-right serial model, that is, with minor or no deterioration in performance in the shield conditions and no compensatory changes in unshielded locations. Further, the serial pathways through the various positions would, in some cases, apparently have to be rather bizarre to produce the observed serial position accuracies.

General discussion

The plan of the remaining discussion will be to designate some characteristics of whole report processing as it takes place in vfs (Fig. 11.1) and with which the present results are consonant. This will be followed by a brief discussion of these characteristics in relation to the present work and a few other particularly pertinent studies. A plausible qualitative overall model of visual whole report processing will then be proposed. Finally, this model, which we call the bounded performance parallel model of whole report vfs activity, will be tested against Rumelhart's multicomponent model in the present data.

Suggested processing characteristics in whole report

1. *Limited capacity.* In performing complete identification of several letters, about four unitary unrelated items (with the complexity of English letters) can be processed without information loss. When display parameters are optimal this processing may include visual identification (via connections with visual LTM); when they are degraded, it will mean the "printing" of the available geometric information in the memory of vfs.

2. *Probabilistic independence.* In the case of unrelated heterogeneous symbols, the probability of one randomly selected item being processed correctly

is independent of whether another item is correctly processed on any single trial. This is predicted by models assuming independence of total completion times (Chapter 4).

3. *Parallelity.* With material that is familiar, but not necessarily orthographically organized, a quantity of items less than or equal to the vfs capacity can be processed in parallel.

4. *Nonuniform capacity allotment.* There may exist an unequal distribution of vfs processing capacity across the items in brief displays. This allotment may reflect reading habits and may, to some degree, be under the control of the observer between trials but not within trials.

The question of capacity will occupy somewhat more space than others, due in part to recent evidence and controversy concerning its limitations. A priori, we may note that capacity limits could, for example, be due to a standard serial type of model with equal processing rates on all items and constant rate across changes in display size or to a limited capacity parallel system (Townsend 1972, 1974b). Note that although the present experiment did not vary the "load" (number of letters), the question of capacity is relevant because of the obvious ceiling on accuracy.

Capacity

When Sperling (1960) found that almost all of the information in the visual array was initially available to be processed, it was natural for him to place the locus of the whole report accuracy limitation in a postvisual short-term (immediate) memory, since the limits of this type of store had long been established. However, in his 1963 and 1967 papers, possibilities of an upper limit in visual scanning capabilities are reasonably manifest. In this context, we may note that the typically suggested upper bound on immediate memory (in afs) is usually assumed to be about seven items, yet a number of studies, particularly with backward masking (like the present), point to an upper bound with visual input of fewer than five items.

Perhaps the best-known mathematical model to place a bottleneck in the visual process is Rumelhart's multicomponent model (1970). In fact, his model appears at first glance just what the present data seem to argue for. It is parallel, limited (in fact, fixed or constant) capacity, and independent, except in partial report, where a reallocation of all capacity to the cued items is supposed to take place. It is also based on exponential probability distributions on completion times, as are many of the results and models considered in the present book. A possible deficit, however, is that this type of model predicts ultimate perfect performance, given a sufficiently long display duration. Nevertheless, only a test of this model against the current data will ascertain whether or not the predicted curves can well approximate the data.

Estes (1972) proposed a qualitative model and reported several experiments that suggest that feature channels interact and effectively yield an early stage of limited capacity in the visual system. There seems little doubt, given Estes' and others' work on homogeneity in visual displays, that interitem similarity

can play an important role is visual processing. However, it appears that under some circumstances, the similarity of distractor items to targets may not play a strong role in recognition (Townsend & Snodgrass 1977). In this latter study, when only two letters were presented close to the center of the fovea, no uniformly degrading effects of similarity of distractor to target were found. Gilmore (1975) presents similar findings. On the other hand, Bjork and Murray (1977) offer convincing evidence that under certain conditions target–distractor similarity is an important variable. More work will have to be done to sort out these contrasting effects.

It appears that there are two studies that have provided much of the basis of arguments to the effect that the initial visual stage is unlimited capacity. These are Eriksen and Spencer (1969) and Shiffrin and Gardner (1972) (see also the first part of this chapter). Both studies involved the comparison of conditions where the items were presented at a very fast rate (simultaneously in the case of Shiffrin and Gardner) vs. a condition where a very slow sequential rate was employed. For example, in the Shiffrin and Gardner experiment, one condition exposed the four items simultaneously for 50 msec and the other exposed them sequentially for 50 msec each. The simultaneously exposed items were as accurately processed as the sequential, if not more so. Gardner (1973) and Shiffrin and Geissler (1973) discuss the theoretical aspects of these and other findings in detail, the latter presenting a simulation model for detection types of paradigms.

The Shiffrin and Geissler model is probably overall the most detailed account of multisymbol visual recognition now available. Although it was not applied to whole report studies, it should not be difficult to adapt it to this purpose. Their approach places the deleterious effect of larger display sizes (assuming sensory factors are held constant) within a higher decisional "interpretation" stage, where items may be confused. A pertinent aspect here, though, is that it *is* accepted that brief exposure, noise masking, lateral interference, and the like can degrade performance, so that the system is not assumed to be perfect. But variation in the number of items to be processed per se is not assumed to affect the *sensory* processing of an individual item.

On the other hand, both the Shiffrin and Gardner (1972) and Eriksen and Spencer (1969) studies, as well as many of the studies by Estes and his colleagues (e.g., Estes & Taylor 1964) were of a detection variety. It seems very likely that in this type of task observers are able to hold the necessary visual information (e.g., a set of critical features or other geometric information) in a visual form system (vfs; see Fig. 11.1) and that this requires much less capacity than the full identification of as many items as possible, as is necessary in visual whole report. It may also be that performance in detection tasks is relatively impervious to the number of items in the display. Thus, the generalizability of detection paradigms to the whole report paradigm may be limited in some respects.

A more recent paper by Wolford and Hollingsworth (1974) provides more direct evidence that visual processing *is* of limited capacity in whole report paradigms. Their finding was that confusion errors were predominantly of a

visual nature but that acoustic confusions, which should occur if the major losses took place in short-term memory, were almost nonexistent. This particular analysis turned out not to be helpful when attempted in the present study, perhaps because of the poststimulus noise mask.

On first examination, the Wolford and Hollingsworth finding might be taken as compatible with the Gardner–Shiffrin idea of postsensory confusability. This cannot be entirely ruled out. However, it seems unlikely that the multinomial type of increase in errors associated with higher-order confusability (Gardner 1973) could lead to the abrupt leveling out at an upper asymptote of about four or four-and-one-half items that occurs in the present type of study. Rather, a limitation in a sensory attentional kind of capacity seems more compatible with these results.

Several papers by Scheerer (1972, 1973, 1974) may be interpreted as suggesting that the inferiority of whole relative to partial report is not due to lack of storage space or defects in report from short-term memory. The shape of the accuracy curves reported by Scheerer (1973: 99) as a function of serial position suggest that the information losses result from a combination of retinal locus, lateral interference, and perhaps a serial or nonuniform parallel processing factor, the latter accounting for the left-to-right decrease in performance.

One of Scheerer's (1974) conditions, the alphabetic condition, comes close to being exactly what is needed to measure the vfs capacity without putting a strain on short-term memory in afs. In this condition the observer has to ascertain which of six critical letters is first in alphabetic order in an array of eight letters. We might expect performance to be slightly better here than we would expect in the true limit because it may have been possible to reject noncritical letters with less than full identification processing, and there were clear constraints that may have been helpful to the observer. For example, the critical letters were *B*, *C*, *D*, *F*, *G*, *H*, and if a *B* were located in the array, then with probability 1 it had to be the correct response. For present purposes, it would be preferred to run the same type of design with all displayed items potentially first alphabetically (the observers also had to report item position for reasons related to that study). In any case, the estimate there of the vfs capacity (taken from his Fig. 3, which did not score for the position judgments) is about five items. With the constraints just mentioned and the fact that no poststimulus mask was employed, this number approximates rather well the four to four-and-one-half items designated in the present study.

There are several aspects of the present study that are compatible with a limited capacity vfs. First, the upper bound on whole report performance has once again been found to be only about four items, perhaps less than the capacity in afs. Second, the shield conditions caused lowered accuracy even though all parts of the display were apparently on sufficiently long to allow establishment of adequate visual images. This is most compatible with a limitation in the visual processing capacity and with the inability to optimally allocate that capacity within trials. Third, the results on interitem indepen-

dence have intuitive, although not direct, implications for where the critical information decrements occur. Independence seems more parsimoniously aligned with limited capacity (parallel) vfs processing inaccuracies than with afs losses.

An increasing appreciation of the influence of lateral interference should be mentioned. It has been found that the negative mutual influences of letters upon one another and the way these interact with retinal locus and number of items displayed are quite robust from very brief to very prolonged viewing time (Townsend et al. 1971; Estes et al. 1976). In these particular studies, as many letters as possible had to be reported, so complete identification was required, just as in whole report studies. Obviously, neither limitations in visual scanning nor short-term memory suffice to explain the poor performance under the elongated viewing intervals. It seems exceedingly likely that these effects occur both in whole report and detection paradigms but that when little discriminative information is needed, as in a detection task, the lateral interference may not manifest itself as a function of number of items. The U-shaped serial position accuracy curves found in many visual studies and a number of other effects show up with very long displays, so they may perhaps be most parsimoniously explained by lateral interferences. However, it is not improbable that some of these effects could also be augmented by other higher-order mechanisms (e.g., serial scanning direction or parallel attention gradients).

We finally point out studies by Klemmer (1964) and Pollack and Johnson (1965) in which binary sequences were briefly exposed to observers taught to use an octal coding technique. Although the short-term memory load was thus reduced by the coding instructions, performance was strongly limited by the number of visual inputs. Although these results also support a limited capacity vfs, there may have been some retinal and lateral interference confoundings due to the large list sizes.

Independence

The present study appears to be the first to analyze statistical dependence in the visual whole report design, although independence has been suggested for other paradigms (Wolford, Wessel, & Estes 1968). We have seen that a model that assumes the letter in each stimulus position is correctly processed independently of the others and that is imbued with guessing structure provides a much better approximation to the data than dependence models, all of which possess reasonable interpretations in terms of possible psychological mechanisms. For instance, the fixed-sample-size model could be used to represent either a constant number of free visual channels or a fixed-size short-term memory buffer (in afs, Fig. 11.1) or even a deterministic constant capacity translator. The Poisson model is most easily identified with a probabilistic serial processor at the visual or translator level, although the random-path Poisson model can be mimicked by a nonindependent parallel model with intratrial capacity reallocation potentialities.

An independence model seems more intuitively compatible with a set of visual channels than with a limited capacity afs or short-term memory. Undoubtedly, though, storage models with the property of independent losses could be constructed, although they might possess somewhat less intuitive appeal. An interesting and perhaps critical test would be an independence analysis of acoustically presented short-term memory items. A discovery of independence there would at least indicate the possibility that the visual whole report losses occur independently in short-term memory, but a finding of, say, a negative correlation would aid greatly in ruling out the afs level as the first major limited capacity (sub)system in visual whole report. Despite the inherent interest of the independence vs. dependence issue as an important aspect of immediate memory storage (in afs), pertinent analyses do not seem to have been reported in the literature.

Parallelity and allotment of capacity

It is convenient to discuss together parallelity and the decreasing left-to-right accuracy gradient. Many investigators have interpreted the accuracy gradient in terms of a serial scanning mechanism operating at some level. It may be, however, that the earlier processing stages operate in parallel and that with several items and a brief exposure, much information is lost during these stages. Certainly there exists some evidence that letter detection can occur in parallel with respect to the visual display (van der Heijden 1975). As noted, the Rumelhart (1970) model is an example of a parallel model applied to whole report activity.

We shall here confine our attention on this question to several pertinent features of the present study. First of all, parallel processing is intuitively compatible with independence, although serial models can be designed that somewhat less intuitively produce this quality also (Townsend 1974b; see also Chapter 4). Next, the accuracy decrement occurring in the shield conditions argues against strict left-to-right processing despite the left-to-right performance asymmetry in the points for any fixed exposure interval.

Further, the nature of the performance alterations in the shield conditions seem not to be qualitatively in line with what one would expect with a semi-random serial process (not entirely left-to-right but with tendencies in that direction). For example, consider observer 2. Scrutiny of the NS condition would suggest an average serial scanning path of the following nature. First: position 1; next: position 2 or 3; next: the other of 2 or 3; next: position 4 or 5; and finally the remaining position. But now in condition LS, the observed changes in accuracy at positions 4 and 5 would most probably be due to alterations of the processing path, because the display time of these two positions was the same as in NS. Here, the average scanning path would presumably be: First: position 1; next: position 5; next: position 2; next: position 3; last: position 4. This seems quite a lot of jumping about to expect of a serial visual scanner.

Even if vfs processing is primarily parallel in the present type of study, however, it seems not unlikely that serial scanning may occur during prolonged iconic storage and especially during translation to afs.

A test of two plausible mathematical models

An overall model that seems compatible with the above conclusions can now be outlined. It is first assumed that any given display arrayed in some retinal pattern has associated with it an upper bound on the visual clarity of the various items within it, which is due primarily to the physical characteristics of the items (the font of an array of letters, the contrast, etc.), to their retinal loci, and to the effects of lateral interference.

Superimposed on this *pattern of retinal integrity* is a visual attention distribution that is usually uniform but may be quite nonuniform, especially where well-ingrained habits are influential (as in reading directions). In cases where highly unfamiliar material is employed, the attention may possibly be focused on fewer items, but is difficult to refocus during a single brief exposure. This attention cannot, however, completely compensate for a low upper bound on the physical integrity, but lack of attention can make accuracy on an item worse.

It is supposed that the items to which attention is directed have part or all of their information printed in a short-term visual memory. When the information from a very well-known item is quite high, visual identification of this item is probably almost automatic. If little geometric information is accrued, then this is held until a decision can be made as to what symbol it represents.

The visual attentional capacity seems to be sufficient to allow about four-and-one-half items to be processed if they are displayed foveally with reasonably high contrast and without severe lateral interference for between 50 and 250 msec, at least when a strong noise mask (or very bright flash mask) cessates the visual image. A lasting icon with a high-integrity display may allow even more items to be correctly reported (perhaps up to the afs capacity).

A mathematical formulation that encompasses the major characteristics in the above account, which we call the *bounded performance model,* was suggested by Townsend and Fial (1968). Let $I(t)$ represent the *integrity* or quantity of stimulus information available at time t for any stimulus location i, and suppose that I grows monotonically with stimulus duration. Let C be a term standing for the total capacity or energy available for devotion to the letter identification task fixed for a given display size, and assume that this capacity is spread over the display according to a distribution $0 \leqslant a_i \leqslant 1$, where $\sum_{i=1}^{5} a_i = 1$, so that a_i is the proportion of attention given to stimulus location i. Now C can be interpreted as the amount of letter information that can be processed via the vfs–LTM system.

It seems reasonable as a first approximation to suppose that the growth of the available stimulus information I is regulated by the simple first-order linear differential equation

$$\frac{dI(t)}{dt} = [I - I(t)]V$$

where V expresses the rate of growth, and $I = \lim_{t \to \infty} I(t)$. That is, the rate of increment to the information quantity $I(t)$ is proportional to the difference between the current level $I(t)$ and the asymptotic level. Integration yields the time function

$$I(t) = I[1 - \exp[-(t - t_0)V]]$$

where t_0 is the time at which information begins to be processed. The probability correct $P(C_i \mid t) = P_i(t)$ via nonguessing at time t is presumed to be equal to the proportion of information accrued on the letter in position i by that time:

$$P_i(t) = \min\left[\frac{I(t)}{I} \times a_i \times C, 1\right]$$
$$= \min(a_i C[1 - \exp[-(t - t_0)V]], 1), \quad 0 < t < +\infty$$

When $a_i C > 1$ – for instance, at long stimulus durations – more attention will be devoted to position i than is necessary for perfect perception; that extra capacity

$$1 - a_i C[1 - \exp[-(t - t_0)V]]$$

will be wasted. The other apparently untoward fact of interest here is that when $a_i C > 1$, even when the product

$$a_i C[1 - \exp[-(t - t_0)V]]$$

is less than 1, the implication is that the observer can "get more out of the proximal stimulus (represented by $I(t)/I = 1 - \exp[-(t - t_0)V]$) than is actually available. A not entirely bizarre interpretation of this state of affairs is that when considerable attention is placed on a developing image, some sophisticated inferential powers may be activated that raise the probability correct over the actual proportion of information processed.

The multicomponent model of Rumelhart (1970) also seems generally compatible with the principles of the above schema. However, the limited capacity is expressed there in terms of a Poisson processing rate (expressed as r below) with the attention allocated to each letter given by some proportion of that rate. As noted earlier, we hypothesized that this model would not be as successful as the bounded performance model because it predicts that the $P_i(t)$ curves, as t varies, increase monotonically to 1, whereas the data appear to evidence a substantial flattening out, within the durations permitted. The bounded performance model does predict this qualitative result.

The appropriate formula for the multicomponent model is

$$P_i(t) = \sum_{j=K}^{\infty} \frac{[\theta_i r(t + 1/\mu)]^j}{j!} \exp[-\theta_i r(t + 1/\mu)]$$

Table 11.2. *Chi-square fits and parameter estimates of the bounded performance model and multicomponent model to the serial position curves alone*

		Parameter estimates							
Observer	χ^2	C	t_0	V	a_1	a_2	a_3	a_4	a_5
Bounded performance model									
1	22.7	4.78	10.0	.068	.30	.19	.19	.13	.19
2	58.9	4.66	10.0	.040	.26	.21	.20	.13	.20
3	32.8	4.48	10.0	.063	.27	.22	.22	.14	.15
4	26.1	4.61	10.0	.058	.28	.21	.21	.14	.16
5	19.9	4.71	10.0	.087	.29	.20	.20	.13	.18
6	31.3	4.23	10.0	.041	.38	.20	.17	.11	.14
7	23.3	4.44	10.0	.074	.28	.21	.19	.14	.18
8	34.9	4.65	10.0	.045	.37	.20	.17	.11	.15
Observer	χ^2	K	$1/\mu$	r	θ_1	θ_2	θ_3	θ_4	θ_5
Multicomponent model									
1	104.8	2	5.25	.458	.32	.20	.21	.08	.19
2	101.6	2	10.00	.382	.28	.23	.20	.09	.20
3	122.1	2	3.70	.380	.29	.22	.27	.12	.11
4	88.1	2	5.46	.416	.29	.15	.24	.09	.13
5	88.6	2	7.94	.697	.32	.21	.29	.06	.12
6	185.5	2	0.0	.220	.43	.21	.17	.08	.12
7	130.4	2	1.64	.360	.30	.20	.20	.11	.18
8	157.5	2	0.0	.254	.41	.21	.15	.08	.15

$\chi^2_{.975}(75) = 100.8, \quad \chi^2_{.80}(75) = 85.1$

where θ_i = proportion of attention allocated to stimulus position i, r = stimulus clarity factor, μ = time constant of decay, and K = criterion number of features that must be completed to identify a (any) letter. Obviously both models implicitly predict independence of processing, before guessing occurs, and therefore are special cases of the general independence model introduced earlier. Each model possesses 7 parameters, against the 75 data points per observer of the basic serial position functions of Phase 1. The models were given the requisite guessing structure and then fit to the Phase 1 serial position data. The parameter estimate of t_0 that optimized the fit of the bounded performance model was between 9 and 10 msec for all observers and so was set to the minimal plausible value of 10 msec in the final version. The latter fits were virtually identical to the optimal fits.

Table 11.2 exhibits the χ^2 values and parameter estimates for each observer for each model. It can be seen that the bounded performance model is substantially superior to the multicomponent model. In fact, the bounded performance model fit is excellent. This can be seen by noting that for large df, the square root of χ^2, χ, is well approximated by

$$\chi \simeq [z + (2df - 1)^{1/2}] \frac{1}{(2)^{1/2}}$$

where z is the unit normal variate; thus the mean or expectation of χ is $E(\chi) \simeq (df)^{1/2} \simeq 8.7$ in the present case and the standard deviation is $\sigma_\chi = .71$. Since the largest value of χ here is $(58.9)^{1/2} = 7.7$, for observer 2, it follows that the bounded performance model is accounting very well for the serial position functions. On the other hand, the multicomponent model fails for all observers at the $\alpha = .20$ level and for six of the eight observers at the .025 level. The χ^2 values for the multicomponent model are typically several times larger than those for the bounded performance model.

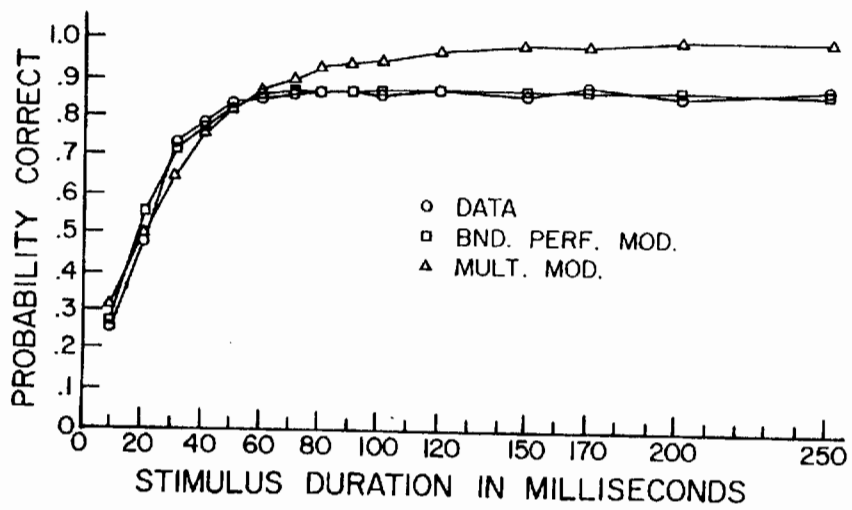


Fig. 11.10. Averaged probability correct as a function of display time compared with the bounded performance and multicomponent models.

The average-number-correct curves (averaged over serial position and subjects) for the two models are shown in Fig. 11.10 and confirm the earlier-stated hypothesis that the bounded performance model would be better able to handle the upper limits on accuracy evident in Fig. 11.6a–i.

The parameter estimates in Table 11.2 indicate that in all cases the capacity C, in terms of number of letters, attained reasonable values of less than 5, although this parameter was unconstrained in the fit program. The consistency of the parameter estimates across the obervers is also striking. The fact that t_0 turned out to be about 10 msec suggests that information begins to be accumulated at the shortest display duration.

It may be noted that in some cases, $aC > 1$; for example, observer 8 exhibited

Table 11.3. *Chi-square fits and parameter estimates of the bounded performance model fit to the combined shielded data and serial position curves*

Observer	χ^2	C	t_0	V	a_1	a_2	a_3	a_4	a_5
1	45.3	6.25	10.0	.074	.47	.15	.14	.10	.14
2	97.7	4.77	10.0	.042	.28	.21	.19	.13	.19
3	89.7	5.19	10.0	.106	.39	.18	.18	.12	.13
4	89.2	6.05	10.0	.094	.46	.16	.15	.11	.12
5	47.3	5.93	10.0	.144	.45	.15	.16	.10	.14
6	96.0	4.18	10.0	.066	.39	.19	.17	.11	.14
7	76.6	4.94	10.0	1.714	.39	.18	.16	.12	.15
8	73.8	4.74	10.0	.062	.40	.19	.16	.11	.14

$\chi^2_{.90}(91) = 108.9, \quad \chi^2_{.80}(91) = 102.0$

$a_1 C = .37 \times 4.6 = 1.7$. As stated earlier, we believe this to represent an excess of capacity devoted to such positions. When t is small, this benefits the favored position, but when t is large, the extra capacity could be more efficiently employed on other positions j, where $P_j(t)$ is still less than 1.

Due to the basically good fit of the bounded performance model on the serial position curves, it was next fit to that data combined with the Phase 2 shield data of Table 11.1. No additional parameters were estimated and t_0 was set to 10 msec. The df was raised to 91 and resulted in the predictions given by THE in Table 11.1. The overall fit characteristics are shown in Table 11.3. They are again statistically excellent for all of the observers. Scrutiny of the shield predictions, however, reveal that the model, while reflecting the overall characteristics, had a tendency to overpredict the amount of decrement in accuracy under the shield condition. There are several factors that appear to have contributed to this tendency. One is the shifting of attention between conditions, resorted to by several observers and sometimes compensating for the increased difficulty on a shielded position. Secondly, as noted earlier, the LS may have permitted a somewhat extended icon that aided performance on the leftmost positions. Thirdly, the model seemed to simply underpredict the control NS accuracy in some cases, especially in positions 4 and 5, which would have established too low a baseline. This latter problem could have been caused by learning effects from Phase 1 to Phase 2 or by an ability (not represented in the model as it stands) to come up with additional capacity under adverse circumstances. The model can be extended to encompass these possibilities but would then possess too much flexibility to allow testing with the current data.

How should we interpret C or the individual capacity components $a_i C$? Presently, it is impossible to decide whether $a_i C$ represents a perceptual filtering action on the display positions or whether it reflects the limits on a single-unit independent visual memory channel in vfs (or perhaps both). In

any event, it will hopefully be feasible to develop the model in more detail and alter interpretations if required; but the basic functional form seems to give an excellent approximation to the data.

Finally, the parallel provision of the model seems most cogent. However, certain intriguing hybrid possibilities are suggested by research indicating that an analog (i.e., time- and space-continuous) change of attention of one part of the visual field to another without change of fixation can occur (Shulman, Remington, & McLean 1979). It is not beyond the realm of credulity that an observer might begin a trial with attention (but not eyes) fixated at a certain letter position (in the NS conditions probably on the left) and then send an analog scan over the remaining positions. The information made available in the respective positions, however, would have to be processed in an independent manner, probably after the analog scan had gone on through the other parts of the array. Note that this model could not be represented as a true serial model because one item is not completed before the next is processed. There also still remain the difficulties posed by the Shiffrin and Gardner (1972) type of data. If the analog scanner were not able to make more than a single pass (and could not be modulated to take advantage of the extra time allotted in the sequential presentation), then the Shiffrin–Gardner findings might not be incompatible with this kind of model. Nevertheless, the straight parallel interpretation seems more plausible and parsimonious at present.