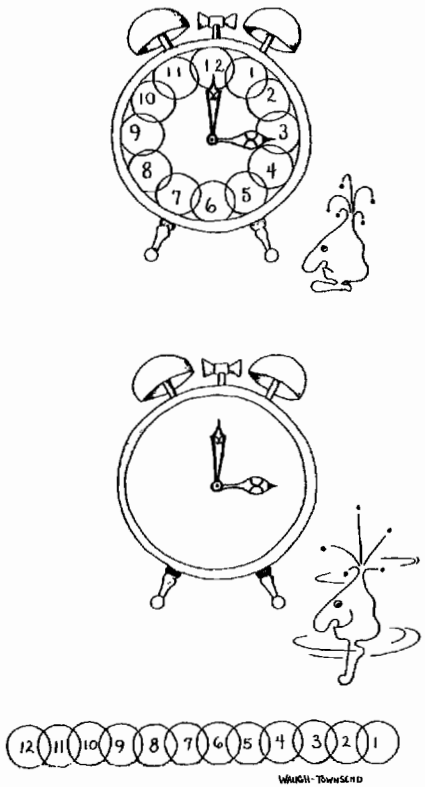


This Durante Creature believes time can be made *additive* if only it can be wrought into a linear mode, which he appears to have accomplished.

12 *Additivity of processing times from separate subsystems and related issues*

Various types of additivity have always played an important role in mathematics, both in its pure and applied forms. In attempting to ascertain the characteristics of a physical system, physicists and engineers typically have at their disposal some type of physical input and output measure on strong scales (e.g., ratio or log interval) and the possibility of some type of additivity expressed in those scales. With a few exceptions, experimental behavioral psychology has had a paucity of such outputs to study; often a single discrete response represents the entire train of internal information processing. Reaction time (RT), which appears to lie on a strong continuous physical scale and to be strategically related to a number of interesting psychological variables,

has been a mainstay in almost all areas of experimental psychology; hence the interest in stochastic latency theory throughout this book.

It was natural for nineteenth-century investigators of psychological phenomena, primarily physiologists and physicists, to turn to RT as one of their tools. Since these early times (roughly 1850 on), there has been a tremendous amount of work accomplished using RT. Much of this work has been oriented toward assumptions that result in average RT being linearly related to certain experimental manipulations.

However, it may help to give some perspective to the present issue to acknowledge the importance of assumptions of additivity in science and accordingly how ubiquitously additivity (or equivalently, linearity) forms one of the base postulates of applied mathematical developments. Although the exact assumptions differ from theory to theory, it is apparent that linearity plays a strategic role in multivariate analysis and analysis of variance, factor analysis, functional measurement (Anderson 1970), and additive conjoint measurement (Krantz, Luce, Suppes, & Tversky 1971). All these methods are important in psychology. Even when we know linearity does not strictly hold, it is often to our advantage to obtain the approximate linear description. For instance, even a simple pendulum is a nonlinear system but may be fruitfully approximated in a linear fashion (Saaty & Bram 1964: 178), and the auditory system is known to evidence nonlinearities (e.g., in the phenomenon known as *combination tones,* where frequencies other than those entering the ear may be heard) but is seemingly well modeled by a linear system at relatively low sound levels. Of course, in some cases it is possible, if often cumbersome, to extend a method or theory to nonlinear domains, as in the instances of nonlinear factor analysis or in polynomial conjoint measurement.

Linearity does not always enter in the same fashion in different theories or methods. Thus, the assumption of additivity of processing times across subsystems in psychology differs markedly from the linearity in the relation of the magnitude of the input to the magnitude of the output that is the standard assumption of linear systems theory, as will be seen below. In fact, we will later suggest that increased application of traditional linear systems theory on the turf of cognitive psychology might be fruitful, particularly as investigators become sufficiently intrepid to hypothesize what is being processed at a certain stage and how much of it, or some transformation of it, is fed into other subsystems. Application of the additive factor method (defined below) and similar techniques may improve our ability and perhaps make us less timorous with regard to the latter, more ambitious undertaking.

In the first major section below, a recently proposed improvement on a traditional RT method of uncovering psychological processes is discussed with regard to serial and parallel systems, and theorems are presented on systems that do or do not yield RT additivity in such methods. In the second section, some concepts from axiomatic, particularly conjoint, measurement theory will be discussed with special reference to the proper measurement scale for

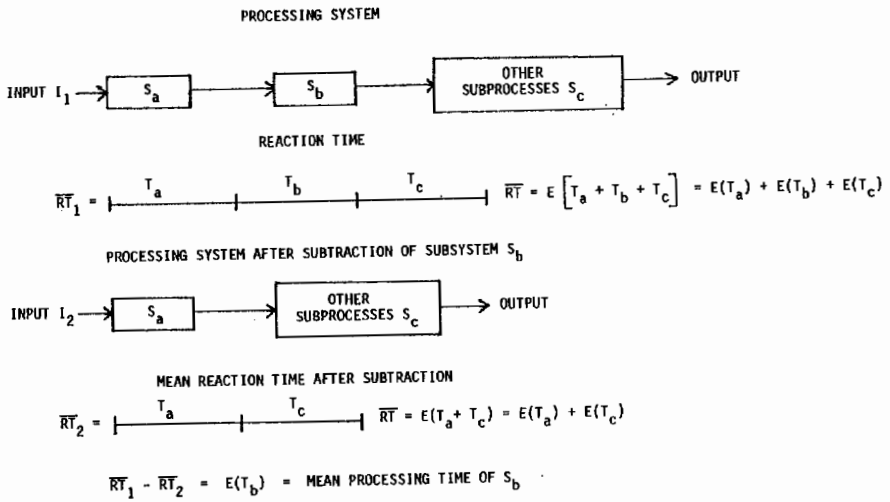


Fig. 12.1. A schematic illustrating the assumptions of Donders' subtractive method.

reaction time and to the related question of whether any transformations of reaction time can induce additivity. A brief consideration of the use of dimensional analysis in psychology will also be taken up here. In the final topic to be considered in this chapter, we shall take a short look at linear systems theory with particular regard to the type of RT additivity treated in the first section. That is, we will ask if there are circumstances where it might make sense to assume linearity both in the input–output magnitudes or amplitudes as well as in the separate processing times. Such questions are also entertained in reference to finite-state automata.

The additive factor method and subsystems arranged serially or in parallel

Every graduate student of experimental psychology learns that an early method of elucidating cognitive processes involved the subtraction of mean RTs of one condition from those of another in order to arrive at the supposed time of a mental operation (see, e.g., Donders 1868, Wundt 1894; see also Fig. 12.1). When mental operations are arranged serially and when changing task demands lead to the complete elimination of an operation or *stage,* we have the notion of a *Donderian system* and a technique known as the *method of subtraction.* This method has never been wholeheartedly accepted, although it is still practiced from time to time. The term *stage* used in this way has been much employed but rarely or never given rigorous mathematical definition in the psychological literature. Recall, however, that in the present work we defined *stage* to demarcate the duration between the successive completions of two elements undergoing processing (see Chapters 1–3). There is a need to

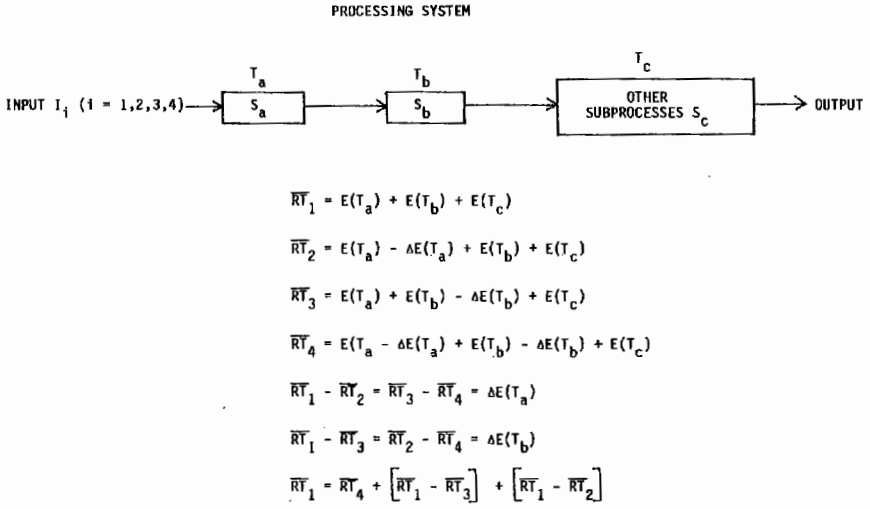


Fig. 12.2. A schematic illustrating the assumptions of Sternberg's additive factor method.

keep the theory regarding the elements being processed clearly distinct from the theory concerning the mechanisms doing the processing. Even though there are many overlaps in the theory and mathematical formalisms associated with the two, some otherwise knotty theoretical dilemmas can be avoided or clarified by maintaining the distinction.

We shall therefore return to our previous terminology where a *system* is a real biological, psychological, or other physical entity, whereas *model* is the mathematical structure presuming to describe the system, make predictions about it, and so on. Instead of a stage being added, deleted, or otherwise affected by experimental procedure, we shall speak of a *subsystem* or *subprocess* as being so affected. However, we wish to first introduce a recently proposed and highly influential method that weakens the assumptions behind Donders' method of subtraction, in the fashion in which it was presented by its author, in order to do as little violence as possible to the original ideas. The approach, developed by Sternberg (1969a, b), is called the *additive factor method,* and is schematized in Fig. 12.2. As noted, it no longer is necessary to postulate that one can insert or take away the contribution of complete stages by experimental manipulations. Rather, it is supposed that experimental factors might be found that, when varied, will affect the mean processing time of separate stages. Figure 12.3a illustrates the additive factor method in a prototypical experimental situation when additivity does hold. In a real experiment corresponding to Fig. 12.3a, the sample mean RTs would approximate the expected RTs shown there. If $\overline{RT}_1 - \overline{RT}_3$ is more or less than $\overline{RT}_2 - \overline{RT}_4$, then an interaction has been found. According to the additive

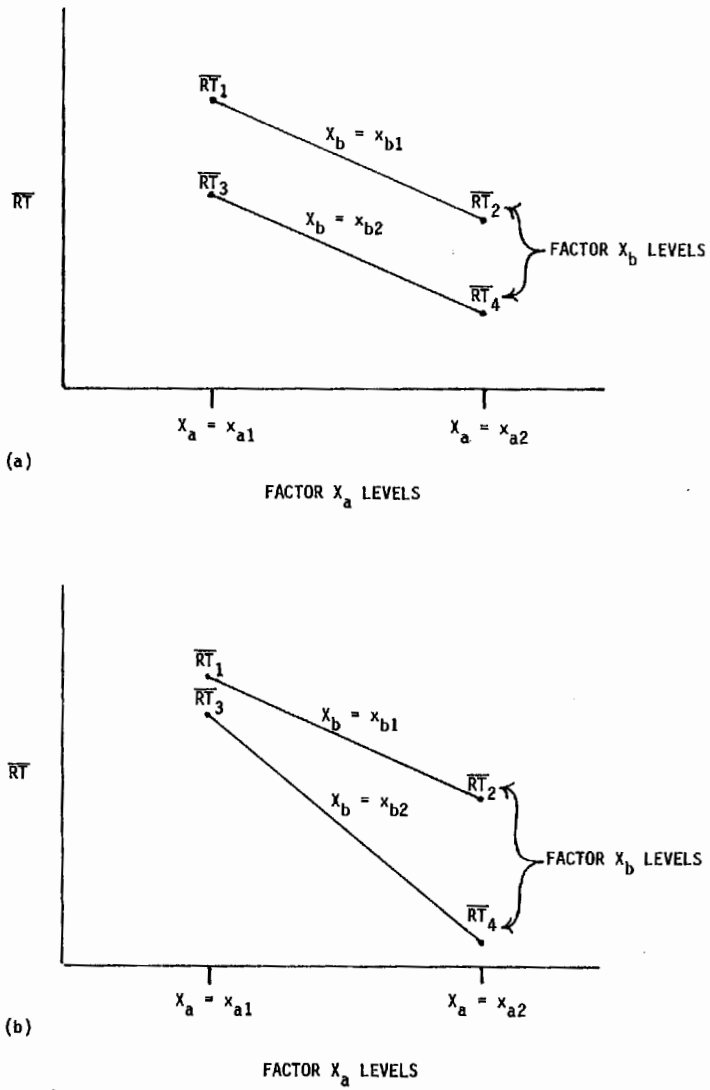


Fig. 12.3. (a) Graph of a hypothetical additive factor experiment with two levels of each factor. Additivity is seen to hold. (b) Graph of a hypothetical additive factor experiment with two levels of each factor. Additivity does *not* hold.

factor method, such an interaction indicates that factors X_a and X_b both affect the same stage, as shown in Fig. 12.3b, and an additivity indicates that X_a and X_b affect different serial stages. Thus, for Sternberg, mean RT additivity should be a fundamental condition for defining a *stage of processing*.

Because experimental factors can often be at least intuitively associated with seemingly separate cognitive operations, the method promises a way of determining which such operations can be segregated into potentially distinct processing stages.

Although Sternberg (1969a, b) postponed a precise definition of *stage* to await further research, he suggested some *potential* conditions that might be incorporated into such a definition:

1. The output or product of a stage should not be affected by factors influencing its duration.
2. The stages in a series should be functionally interesting, qualitatively different, and reasonable in light of other knowledge of psychological behavior.
3. A stage should be able to process only one input or signal at a time.
4. Stage durations should be stochastically independent, i.e., $f_{ab}(t_a, t_b) = f_a(t_a)f_b(t_b)$ in serial systems.

Anticipating our own approach, we do not believe that an interaction should be taken as suggesting a single processing unit (stage, subsystem, or whatever) unless there are other, very strong reasons for doing so. On the other hand, a strong finding of additivity does suggest separate processing operations to us, at least in a systemic sense.

It is our view that mean RT additivity (and at some point various psychologically important types of interactions) should be logically and mathematically investigated with regard to what types of subsystem hookups do or do not violate it. Very little has been published that is relevant to this aim.

A major goal of the present chapter will be to begin such an investigation. However, a brief comment on the ancillary "potential" conditions is in order, centered around the attitude that it is more fruitful to investigate the various facets of coupled interactions of subsystems than to emphasize the defining of a particular type of subsystem (in this case, stage). These remarks are not intended as a criticism of Sternberg's considered usage of conditions 1–4 (particularly since Sternberg's suggestions were only tentative) so much as an attempt to see how they intersect with our approach.

First, condition 1 rules out many interesting subsystems – for example, ones where the integrity of the input affects its processing duration. Number 2 is a reasonable requisite for a psychologically viable subsystem. Requirement 3 must be ruled out as a condition in our approach because it precludes parallel processing by a single subsystem, which there is reason to believe is an important type of functioning in neurological complexes (Grossberg 1978). Finally, condition 4 should not be a requirement, although it certainly renders system analysis more facile. Sternberg (1969a, b) himself did not assume condition 4 in his development of the additive factor method.

We next turn to our own definitions and then demonstrate the most natural (and obvious) class of systems that produce mean RT additivity. The logical independence of additivity and stochastic independence is also demonstrated. Other aspects and potential pitfalls associated with the procedure are mentioned.

The following subsection discusses a class of distribution-free independent parallel models and presents two results indicating the difficulty or impossibility of their predicting mean RT additivity. Next, a characteristic function approach to RT additivity is taken with the result that the potential class of models satisfying the constraint in a strong form is essentially that of stochastically independent serial models. Finally, the models of a class of distribution-free mean-additive serial systems is exhibited and the failure of the class of independent parallel models (mentioned above) to predict mean additivity is related to some parallel–serial nonequivalence theorems. These are then given specific form with the aid of exponential intercompletion time models.

Let us now attempt to be more precise concerning the additive factor method. RT is, as usual, considered to be represented as a random variable in this book.

We adopt the following notation:

S = total (real) system under observation

S_i = subsystem of particular interest, $i=a, b, \ldots$

$\mathbf{RT}$ = overall reaction time random variable

X_i = experimentally manipulable factor, formalized as a set; in particular circumstance, may be a numerical scale of some sort, $i=a, b, \ldots$

$\mathbf{T}$ = random variable representing the time duration of the entire system S from time $t=0$ to completion of all processing

$\mathbf{T}_i$ = random variable representing the duration of serial subsystem S_i, $i=a, b, \ldots$

$\mathbf{T}_i$ = random variable representing the duration of parallel subsystem S_i, $i=a, b, \ldots$

$f_i, F_i, \bar{F}_i$ = density function, distribution function, and survivor function on the $\mathbf{T}_i$, respectively, of subsystems arranged serially, $i=a, b, c, \ldots$

$g_i, G_i, \bar{G}_i$ = density function, distribution function, and survivor function on the $\mathbf{T}_i$, respectively, of subsystems arranged (independently) in parallel, $i=a, b, c, \ldots$

The probability functions (f_i etc.) are assumed to be functions of the X_i sets so that, for instance, we have $f_i(t_i; x_i)$ as the density function on the processing duration consumed by subsystem S_i under the influence of $x_i \in X_i$. When not concerned with a particular value x_i, we shall often just refer to X_i. If, in fact, the two subsystems S_a, S_b are functioning serially and in a stochastically independent fashion, and factor X_i affects only S_i $(i=a, b)$, then the overall density function on total processing duration for both subsystems is the convolution (see, e.g., Chapter 3)

$$f_{ab}(t; X_a, X_b) = f_a(t; X_a) * f_b(t; X_b)$$

Of course, in general, for a given t,

$$f_{ab}(t; X_a, X_b): X_a \times X_b \to \mathbf{R}^+ \quad \text{and} \quad \int_0^\infty f_{ab}(t; X_a, X_b)\, dt = 1$$

that is, f is a mapping from the Cartesian product of the sets X_a, X_b into the positive real numbers with total measure 1, as befits a density function.

It follows immediately that, in this serial case, the overall mean processing time is an additive function of the factors X_a and X_b,

$$E_{ab}(\mathbf{T}; X_a, X_b) = E(\mathbf{T}_a; X_a) + E(\mathbf{T}_b; X_b) \tag{12.1}$$

This result is the basis of the additive factor method. Note, though, that whereas we have shown seriality of the subsystems to be sufficient for mean RT additivity, we have not demonstrated that it is necessary, and in fact, as we will see, it is not. Some parallel models do predict mean RT additivity. However, we will also demonstrate that some quite broad classes of parallel systems do not.

Sternberg (1969a: 280) also stated that (standard) serial short-term memory search (see Chapter 6) can be characterized as a special case of the additive factor method and consequently, so can Donders' subtractive method. A serial scanning model results when $X_a, X_b, X_c, \ldots$ are each such that $X_i = \{0, 1\}$ and

$x_i = 1$ if subsystem i is present
$= 0$ if subsystem i is absent, $i=a, b, c, \ldots$

where subsystem S_i contributes a nonzero duration if and only if $x_i = 1$. In this event, $E(\mathbf{T}; x_a, x_b, \ldots) = E(\mathbf{T}_a; x_a) + E(\mathbf{T}_b; x_b) + \cdots$ and mean RT is additive. Further, one may estimate $E(\mathbf{T}_i; x_i)$ if the x_j can be set to zero for every $j \neq i$, for instance, by task manipulation. If $E(\mathbf{T}_i; 1) = E(\mathbf{T}_j; 1)$ for all i and j, the standard serial model results. If $E(\mathbf{T}_i; 1) = 1/u$, where u is the exponential rate, the standard exponential serial model is produced.

If S_i is a location in a memory storage buffer and $x_i = 1$ or 0 corresponds to the presence or absence of an item in S_i, then in this characterization the search of a buffer location consumes time if and only if there is an item present. Note that in this particular example, a larger value of x_i decreases speed rather than increasing it, as we usually assume in this chapter. We shall bring up the parallel analog of this special case after Proposition 12.4 below.

In most applications of the additive factor method two or more factors are manipulated – for example, X_a and X_b (i.e., select different members from the respective sets) – and then an analysis of variance is performed to test for main effects and for interactions among the factors. If all interactions are nonsignificant, then one concludes that there do indeed exist two separate serial subsystems.

There is something of an inductive asymmetry in the method. If one finds solid evidence for additivity, then support is immediately accrued for serialistic additive sybsystems or for nonserial subsystems that are interdependent in

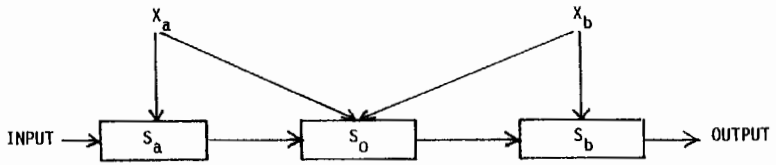


Fig. 12.4. Schematic showing that a two-way interaction could occur via a third stage or operation (S_0). If a third factor X_0 could be found that influences only S_0 and S_a, or only S_0 and S_b, but not all three, no three-way interaction would appear. Thus, the experimenter could conclude that two separate stages S_a and S_b existed.

an additive fashion. On the other hand, if two factors that are thought to affect separate subsystems interact, perhaps one has chosen the wrong factors, and maybe further search will lead to two other factors that are also believed to affect those subsystems but which do not interact. This is something like the classic search for a nonblack crow in an infinite population of crows (assuming the definition of *crow* does not imply blackness): No matter how many crows one inspects, there is always the possibility that the next time one will finally find a crow of a different color.

A related problem is that a single two-factor experiment with factors X_a and X_b may exhibit a two-way interaction because the two factors affect a subsystem in common (S_0) although they also affect two other subsystems (S_a, S_b) separately (see Fig. 12.4). This problem can in principle be solved (if one has reason to suspect its presence) by manipulating a third factor X that does not affect the common subsystem but does affect one or both of the "unique" subsystems. The absence of a three-way interaction suggests the existence of the two unique subsystems, S_a and S_b. Hence, S_a and S_b contribute additively to RT. The factors X_a and X_b do not selectively influence these subsystems, however, and so their additivity is empirically disguised. This is not the only possible conclusion from such a finding, although it does seem a plausible one.

Quite different problems relate to the statistical testing of additivity. Because additivity is usually the null hypothesis, "proving" the null hypothesis is equivalent to supporting separate subsystems (as well as, potentially, certain types of connections, as in serial hookups). The supposition that *any* functional subsystems of the brain are purely additive is a strong claim, so that special insurance should be taken to preclude indiscriminate acceptance of the null hypothesis. Sternberg (1969a) points out this pitfall and recommends several types of insurance – for instance, that investigators employ a Type I error probability equal to that of some reasonable theoretical alternative Type II error likelihood, rather than the typical $\alpha = .05$. However, many papers have appeared where this caveat has not been heeded.

Other difficulties of unknown severity arise from the assumptions in the analysis-of-variance methodology of normality and homogeneity of variance

both of which are obviously violated in typical RT data. Recently, however, Ashby and Townsend (1980) suggested a nonparametric alternative to analysis of variance, based on the RT distributions rather than the means, which can be used in conjunction with the additive factor method, in the same way that analysis of variance can, but which avoids many of its inappropriate assumptions.

We now turn to the problem of summarizing the behavior of serial systems with regard to influences of the experimental factors, X_a and X_b. For simplicity, we continue to assume S_a always processes its input first and that this is followed by S_b working on its input, which is typically the output of S_a. We assume in the following proposition that the first- and second-order partial derivatives exist and are continuous over the positive real line. When we wish to emphasize that the expectation is taken relative to $\mathbf{T}_a$, $\mathbf{T}_b$, or both, we write $E_a(\cdot)$, $E_b(\cdot)$, or $E_{ab}(\cdot)$, respectively.

Proposition 12.1: (i) Mean RT additivity holds if X_a affects only the distribution of processing time on $\mathbf{T}_a$ and X_b affects only the distribution of processing time on $\mathbf{T}_b$, irrespective of $\mathbf{T}_a$, in the sense that

$$f_a(t_a; x_a, x_b) = f_a(t_a; x_a) \tag{12.2a}$$

and

$$\frac{\partial^2}{\partial x_a \partial x_b} \int_0^\infty \int_0^\infty f_a(t_a; x_a) f_b(t_b; x_a, x_b \mid t_a) t_b \, dt_b \, dt_a = 0 \tag{12.2b}$$

(ii) Mean RT additivity *may not* hold if X_a affects the distribution of processing time $\mathbf{T}_b$ or X_b affects the distribution of processing time $\mathbf{T}_a$ in addition to X_a affecting that of $\mathbf{T}_a$ and X_b that of $\mathbf{T}_b$.

Proof: (i) Mean RT additivity holds if

$$\frac{\partial^2}{\partial x_a \partial x_b} E_{ab}(\mathbf{T}_a + \mathbf{T}_b; x_a, x_b) = 0 \tag{12.3}$$

Now

$$\begin{aligned} E_{ab}(\mathbf{T}_a + \mathbf{T}_b; x_a, x_b) &= \int_0^\infty \int_0^\infty f_a(t_a; x_a) f_b(t_b; x_a, x_b \mid t_a)(t_a + t_b) \, dt_b \, dt_a \\ &= E_a(\mathbf{T}_a; x_a) + \int_0^\infty \int_0^\infty f_a(t_a; x_a) f_b(t_b; x_a, x_b \mid t_a) t_b \, dt_b \, dt_a \\ &= E_a(\mathbf{T}_a; X_a) + E_a[E_b(\mathbf{T}_b; X_a, X_b \mid \mathbf{T}_a)] \end{aligned}$$

Taking the second mixed partial derivative produces 0 as a result, which implies the functional representation

$$E_a(\mathbf{T}_a; X_a) = a(X_a)$$

$$E_a[E_b(\mathbf{T}_b; X_a, X_b \mid \mathbf{T}_a)] = b_1(X_a) + b_2(X_b)$$

so that

$$E_{ab}(\mathbf{T}_a+\mathbf{T}_b; X_a, X_b)=a(X_a)+b_1(X_a)+b_2(X_b)$$
$$=a'(X_a)+b'(X_b)$$

are additive functions of X_a and X_b.

(ii) Let

$$f_a(t_a; x_a)=\frac{1}{x_a}\exp[-(t_a/x_a)], \quad 0\leqslant t_a<\infty$$

and

$$f_b(t_b; x_a, x_b \mid t_a)=\frac{1}{x_a x_b}\exp[-t_b/(x_a x_b)], \quad 0\leqslant t_b<\infty$$

Then Eq. 12.2a holds but Eq. 12.2b does not; in particular,

$$\frac{\partial^2}{\partial x_a \partial x_b}\int_0^\infty\int_0^\infty f_a(t_a; x_a) f_b(t_b; x_a, x_b \mid t_a) t_b\, dt_b\, dt_a$$
$$=\frac{\partial^2}{\partial x_a \partial x_b}\int_0^\infty\int_0^\infty \frac{1}{x_a}\exp[-(t_a/x_a)]\frac{1}{x_a x_b}\exp[-t_b/(x_a x_b)] t_b\, dt_b\, dt_a$$
$$=\frac{\partial^2}{\partial x_a \partial x_b}(x_a x_b)=1$$

and in fact,

$$E_{ab}(\mathbf{T}_a+\mathbf{T}_b; x_a, x_b)=x_a+x_a x_b$$

clearly an interactive function of x_a and x_b. □

Strictly, Eq. 12.3 implies linearity so that, in general

$$E_{ab}(\mathbf{T}_a+\mathbf{T}_b; x_a, x_b)=E_a(\mathbf{T}_a; x_a)+E_b(\mathbf{T}_b; x_b)+C$$

where C is a constant, but this in no way affects the outcome because C can be absorbed into $a'(X_a)$ or $b'(X_b)$.

It may be observed that if we postulated that the expectation of $\mathbf{T}_b$ were either a function of X_a alone or a nonadditive function of X_a and X_b, then Eq. 12.3 would imply that

$$E_a[E_b(\mathbf{T}_b; X_a, X_b \mid \mathbf{T}_a)]=b(X_b)$$

A simple example of part (i) of Proposition 12.1 would be to let $f_a(t_a; x_a)$ be as in part (ii) of the above proof, but let

$$f_b(t_b; x_a, x_b \mid t_a)=\frac{1}{x_b}\exp[-(t_b/x_b)]$$

that is, not a function of x_a or t_a.

One intriguing property of the additive factor method is that one may obtain mean RT additivity even though the within-trial time contributions of

separate subsystems are not independent, even in serial systems. The independence we refer to here is the so-called *stochastic independence* and corresponds to what we referred to elsewhere in the book as *across-stage independence*. The two are equivalent because the intercompletion times of subsystems play a formal role analogous to the intercompletion times of elements referred to in earlier chapters. Even though we wish to make the formal similarity evident, we shall use somewhat distinct terminology where it seems appropriate in order to make clear that we are modeling different material situations. Thus, we will refer here to *stochastic independence* between subsystems rather than to *across-stage independence*.

The following proposition clarifies the relationship between stochastic independence and mean RT additivity.

Proposition 12.2: Mean RT additivity is logically independent of *stochastic independence*.

Proof: To prove this we need to exhibit four models: (1) mean RT additive and stochastically independent; (2) mean RT interactive and stochastically independent; (3) mean RT additive but stochastically dependent; (4) mean RT interactive and stochastically dependent. The proof will be carried out within the context of the serial model. Thus, $\mathbf{T}_a$, $\mathbf{T}_b$ are both the intercompletion time random variables as well as the actual processing time random variables for the models of systems S_a and S_b.

For stochastic independence to hold, it is clearly necessary and sufficient that

$$f_{ab}(t_a, t_b; x_a, x_b)=f_a(t_a; x_a, x_b) f_b(t_b; x_a, x_b)$$

Notice that it is not necessary that f_a, f_b be functions only of x_a, x_b, respectively, although that might occur [i.e., $f_a(t_a; x_a, x_b)=f_a(t_a; x_a)$, for example]. Similarly, for mean RT additivity to hold it is necessary and sufficient that

$$E_{ab}(\mathbf{T}; x_a, x_b)=E_{ab}(\mathbf{T}_a+\mathbf{T}_b; x_a, x_b)=a(x_a)+b(x_b)$$

(1) *Mean RT additive and stochastically independent.* Obviously this can readily be obtained by setting

$$f_{ab}(t_a, t_b; x_a, x_b)=f_a(t_a; x_a) f_b(t_b; x_b)$$

Stochastic independence is thereby fulfilled and

$$E_{ab}(\mathbf{T}_a+\mathbf{T}_b; x_a, x_b)=\int_0^\infty\int_0^\infty f_a(t_a; x_a) f_b(t_b; x_b)(t_a+t_b)\, dt_b\, dt_a$$
$$=E_a(\mathbf{T}_a; x_a)+E_b(\mathbf{T}_b; x_b)=a(x_a)+b(x_b)$$

As an example, let $X_i=\mathbf{R}^+$, $i=a, b$. Then,

$$f_a(t_a; x_a)=x_a\exp(-x_a t_a), \quad f_b(t_b; x_b)=x_b\exp(-x_b t_b)$$

where $x_a, x_b \in X_a, X_b$, respectively. Thus, the experimental factors are the rates in the independent serial exponential densities.

(2) *Mean RT interactive and stochastically independent.* This case is produced when, for example,

$$f_{ab}(t_a, t_b; x_a, x_b) = f_a(t_a; x_a) f_b(t_b; x_a, x_b)$$

The stochastic independence is evident and the mean RT interaction comes from

$$E_{ab}(\mathbf{T}_a + \mathbf{T}_b; x_a, x_b) = E_a(\mathbf{T}_a; x_a) + E_b(\mathbf{T}_b; x_a, x_b)$$

as is readily demonstrated by integration where $E_b(\mathbf{T}_b; x_a, x_b)$ does *not* further devolve into an additive function of x_a and x_b. One example was given in the proof of part (ii) of Proposition 12.1. Another similar example is found with $X_i = \mathbf{R}^+$, $i = a, b$, and

$$f_a(t_a; x_a) = x_a \exp(-x_a t_a)$$

$$f_b(t_b; x_a, x_b) = x_a x_b \exp(-x_a x_b t_b)$$

(3) *Mean RT additive and stochastically dependent.* Here we require

$$f_{ab}(t_a, t_b; x_a, x_b) \neq f_a(t_a; x_a) f_b(t_b; x_b)$$

Nevertheless, mean RT additivity still can occur whenever

$$E_{ab}(\mathbf{T}_a + \mathbf{T}_b; x_a, x_b) = \int_0^\infty \int_0^\infty f_{ab}(t_a, t_b; x_a, x_b)(t_a + t_b)\, dt_b\, dt_a$$
$$= E_a(\mathbf{T}_a; x_a) + E_b(\mathbf{T}_b; x_b)$$

However, it may not be obvious that this can in fact occur since X_a affects $\mathbf{T}_a$, which, through the stochastic dependency, affects $\mathbf{T}_b$. Therefore, an example is called for. Suppose in addition to serially arranged S_a, S_b a third system S_c acts as a capacity source affecting the probability densities associated with S_a and S_b. Suppose that the amount of capacity is a random variable $\mathbf{C}$ and that the three probability functions are

$$f_a(t \mid c) = c \exp[-c(t - x_a)], \quad x_a \leqslant t < \infty$$

$$f_b(t \mid c) = c \exp[-c(t - x_b)], \quad x_b \leqslant t < \infty$$

$$P_c(c) = \tfrac{1}{2}, \quad c = 2$$
$$= \tfrac{1}{2}, \quad c = 4$$

Observe that the experimental factors affect subsystems S_a and S_b this time, not by manipulating the rates of processing but rather by shifting up or down the entire processing time distributions. It is probably evident that a positive dependency exists between S_a and S_b whatever $P_c(c)$ because both systems will have high capacity or both will have low capacity on any one trial. Nevertheless, to make matters crystal clear, we carry out the example with the

$$f_{ab}(t_a, t_b; x_a, x_b) = P(c=2) f_a(t \mid c=2) f_b(t \mid c=2)$$
$$+ P(c=4) f_a(t \mid c=4) f_b(t \mid c=4)$$
$$= \tfrac{1}{2}\{2\exp[-2(t_a - x_a)]\, 2\exp[-2(t_b - x_b)]$$
$$+ 4\exp[-4(t_a - x_a)]\, 4\exp[-4(t_b - x_b)]\}$$

Note from the above that given a particular value of c, the system is stochastically independent, but that averaging over c produces dependence as we show below. First observe that the two marginal densities are

$$f_a(t_a; x_a) = \int_{x_b}^\infty f_{ab}(t_a, t_b; x_a, x_b)\, dt_b$$
$$= \tfrac{1}{2}\{2\exp[-2(t_a - x_a)] + 4\exp[-4(t_a - x_a)]\}$$

and

$$f_b(t_b; x_b) = \tfrac{1}{2}\{2\exp[-2(t_b - x_b)] + 4\exp[-4(t_b - x_b)]\}$$

Now, one natural way to demonstrate that the dependency is really a positive one is to compute the probability that S_b completes its task in some time greater than t_b *given* that S_a takes time $t > t_a$ to finish its task (e.g., see Chapter 4). This may be compared with the marginal probability that S_b takes longer than t_b to do its job. If the dependence is positive, the conditional probability will be greater than the marginal (i.e., longer times for S_a imply longer times for S_b and conversely).

Thus, we wish to compute

$$P(\mathbf{T}_b > t_b \mid \mathbf{T}_a > t_a) \text{ vs. } P(\mathbf{T}_b > t_b), \quad x_a < t_a; \quad x_b < t_b$$

which can be expressed as

$$P(\mathbf{T}_b > t_b \mid \mathbf{T}_a > t_a) = \frac{P(\mathbf{T}_b > t_b \cap \mathbf{T}_a > t_a)}{P(\mathbf{T}_a > t_a)}$$
$$= \frac{\tfrac{1}{2}\{\exp[-2(t_a - x_a)]\exp[-2(t_b - x_b)] + \exp[-4(t_a - x_a)]\exp[-4(t_b - x_b)]\}}{\tfrac{1}{2}\{\exp[-2(t_a - x_a)] + \exp[-4(t_a - x_a)]\}}$$
$$\text{vs.} \quad \tfrac{1}{2}\{\exp[-2(t_b - x_b)] + \exp[-4(t_b - x_b)]\}$$

With a little effort this becomes

$$\{1 - \exp[-2(t_b - x_b)]\}\{1 - \exp[-2(t_a - x_a)]\} \quad \text{vs.} \quad 0$$

and because the left-hand side is always positive for $t_a > x_a$, $t_b > x_b$, we realize the dependency is indeed of a positive quality. That is, knowing that S_a took longer than some specified duration increases the likelihood that S_b took longer than some other arbitrary duration.

$$E(\mathbf{T}_a+\mathbf{T}_b; x_a, x_b)=E_c[E_{ab}(\mathbf{T}_a+\mathbf{T}_b; x_a, x_b \mid \mathbf{C})]$$
$$=E_c\left(\frac{1}{\mathbf{C}}+x_a+\frac{1}{\mathbf{C}}+x_b\right)$$
$$=x_a+x_b+\tfrac{1}{2}(\tfrac{2}{2}+\tfrac{2}{4})$$
$$=x_a+x_b+\tfrac{3}{4}$$

which is a clearly additive function of the experimental factors X_a, X_b, as claimed.

(4) *Mean RT interactive and stochastically dependent.* This time we instill a direct positive dependence, by letting the exponential rate of S_b be inversely proportional to the duration of processing consumed by S_a:

$$f_a(t_a; x_a)=\tfrac{1}{2} \quad \text{if } t_a=x_a$$
$$=\tfrac{1}{2} \quad \text{if } t_a=Mx_a, \quad 0<M, \quad x_a<+\infty$$

and

$$f_b(t_b; x_b \mid t_a)=\frac{x_b}{t_a}\exp[-[(x_b/t_a)t_b]], \quad 0<x_b<+\infty$$

The f_a probability mass function is contrived; more complicated functions could be used, but the present will suffice to illustrate the case.

The positive stochastic dependency is so transparent here that we will not bother to compute more statistics. However, we still need to exhibit the mean RT,

$$E(\mathbf{T}; x_a, x_b)=E(\mathbf{T}_a; x_a, x_b)+E_a[E_b(\mathbf{T}_b; x_a, x_b \mid \mathbf{T}_a=t_a)]$$
$$=E_a\left(\mathbf{T}_a+\frac{\mathbf{T}_a}{x_b}\right)=\frac{1}{2}\left(x_a+\frac{x_a}{x_b}\right)+\frac{1}{2}\left(Mx_a+\frac{Mx_a}{x_b}\right)$$
$$=\frac{M+1}{2}x_a\left(1+\frac{1}{x_b}\right)$$

which is patently interactive. Notice that X_a increases the processing time of S_a, whereas X_b decreases the processing duration of S_b and that when X_b becomes larger it partially but not completely offsets the effects of X_a (due to the interaction). □

Thus, we see that mean RT additivity and stochastic independence are logically independent. Nevertheless, intuitively it seems that when a stochastic dependence between two serially arranged subsystems is caused by an influence from S_a to S_b, then mean RT is likely to be interactive, and vice versa. Such might be the case when the nature of the output from S_a affects the time $\mathbf{T}_b$ of S_b, and X_a in turn influences the time and output of S_a.

Taylor (1976c) suggested an extension of the additive factor method built on the concept of stages that have at most a dependence of a linear sort. One

consequence of Taylor's results is the encompassing of a class of time-overlapping subprocesses within the equivalence class of systems supported by the finding of factor additivity. For instance, suppose subsystem S_a and subsystem S_b always overlap in their processing times by an amount k. Then total RT is $\mathbf{RT}=\mathbf{T}_a+\mathbf{T}_b-k$, where $\mathbf{T}_a$ and $\mathbf{T}_b$ are the respective processing times of S_a and S_b. The important point is that if factor X_a affects only S_a and factor X_b affects only S_b, then $E(\mathbf{RT})=E_a(\mathbf{T}_a; x_a)+E_b(\mathbf{T}_b; x_b)-k$; that is, we will observe a mean-additive function of the two subprocesses. The mathematical annoyances of what happens when $\mathbf{T}_b<k$ and the like have not been investigated. One way out would be to put a bound on the distribution of $\mathbf{T}_b$ [$P(\mathbf{T}_b \leqslant k)=0$], but this approach may appear somewhat artificial. Although the class of linearly dependent processes proposed by Taylor (1976c) includes many more models than that exhibited just above, we suspect that stochastic parallel models will not typically fall into this class. Readers interested in further pursuing the additive factor method and related techniques may wish to consult methodological papers by Pachella (1974), Taylor (1976c), McClelland (1979), Schweickert (1978), and Ashby and Townsend (1980).

Along different lines, Schweickert (1978) developed an interesting method based on critical path analysis that should expand the purview of subsystem-connection investigation. This method was briefly described in Chapter 2.

From a modeling point of view, perhaps the most important aspect of the additive factor method is its applicability to a wide variety of experimental circumstances together with the fact that reasonably large classes of models predict additivity whereas many other models predict interactions. We shall see below that when we tie certain aspects of the experiment (e.g., various characteristics of the stimulus and response) to structure in the models, we greatly enhance our testability potential. This idea will be elaborated further in succeeding chapters when discussing experimental and mathematical discrimination of parallel and serial processes. It is not a new idea, of course, but only since the advent of reasonably elegant mathematical models in psychology has such cross-situational testing really meant very much. The reason is that highly verbalized theories (Tolman 1932) are easy to bandy about, and inelegant mathematical theories (Hull 1943) also permit theory-saving stratagems that may be somewhat self-serving. Estes (1959) discusses such matters within the context of learning theory.

Additivity of processing times in parallel systems

We focus our discussion in this section on two subsystems S_a and S_b, and assume that experimental factors can be found that selectively affect S_a and S_b. As before, call X_a the factor that affects S_a, and X_b the factor that affects only S_b. In order to keep this discussion within bounds it will be assumed that X_a and X_b are ordered sets that, unless otherwise noted, tend to affect RT by decreasing it (increasing speed) as X_a or X_b increases. Let the index j on x_{aj}, $j=1, 2, \ldots$, represent distinct experimenter-controlled values

of X_a and similarly for x_{bj}. Assume that $x_{aj} > x_{ai}$ if $j > i$. We wish to investigate the ability of parallel systems to produce additivity of processing times. It might be anticipated that a complete answer would follow immediately from our earlier results on parallel-serial equivalence. However, such is not the case.

The reason that the present conclusions differ somewhat is not because of our focus on subsystems rather than elements, for many of the preceding and following results apply in either case. The conclusions differ because we are now asking about what is entailed when specific experimental factors are manipulated that presumably affect separate subsystems or elements. Logically, it is irrelevant whether we ask this about subsystems or about elements that may be features, gestalts, or whatever.

We begin our investigations with a parallel analog to Proposition 12.1. The analog is not quite as powerful as Proposition 12.1; nevertheless, it covers a very broad class of parallel models.

Because we are not necessarily going to assume independence of (total) completion times of the individual parallel subsystems at the present time, we will have to express the various functions in terms of both X_a and X_b. Also, it will be convenient to use the total completion times. Recall that the total completion time of an element (i.e., subsystem here) associated with position a is defined by the sum of the intercompletion times up to a's point of completion. Thus, if n_a is the number of subsystems that complete processing up to and including the instant of a's completion, then we define the random variable $\mathbf{T}_a = \sum_{i=1}^{n_a} \mathbf{T}_i$, where $\mathbf{T}_i$ is a random variable representing the ith intercompletion time. Of course, in parallel operations, $\mathbf{T}_i$ is also the actual random processing time of the element in position i. In general, a dependence of S_a on X_b or S_b on X_a might occur indirectly through the potential dependency inherent in

$$P_{ab}(\mathbf{T}_a \leqslant \tau_a \cap \mathbf{T}_b \leqslant \tau_b; x_a, x_b)$$

Because of this possible indirect dependence, even the marginals

$$P_i(\mathbf{T}_i \leqslant \tau_i; x_a, x_b), \quad i = a, b$$

must retain a *potential* functional dependence on both X_a and X_b.

Let

$$P_{ab}(\mathbf{T}_a \leqslant \tau \cap \mathbf{T}_b \leqslant \tau; x_a, x_b) = G_{ab}(\tau; x_a, x_b)$$
$$= \text{distribution function on } \max\{\mathbf{T}_a, \mathbf{T}_b\}$$

Let us define superadditivity as being satisfied if, when $x_{i1} < x_{i2}$ $(i = a, b)$, then

$$E[\max(\mathbf{T}_a, \mathbf{T}_b); x_{a2}, x_{b1}] - E[\max(\mathbf{T}_a, \mathbf{T}_b); x_{a2}, x_{b2}]$$
$$< E[\max(\mathbf{T}_a, \mathbf{T}_b); x_{a1}, x_{b1}] - E[\max(\mathbf{T}_a, \mathbf{T}_b); x_{a1}, x_{b2}]$$

that is, the smaller the X_a factor is, the bigger effect a given X_b decrease has in terms of increasing RT. Subadditivity is defined similarly, but in reverse.

The reader may wish to reexamine the earlier mean RT interactive models to determine whether they reveal sub- or superadditivity (if either).

Proposition 12.3 provides a large class of parallel models that cannot predict mean RT additivity. It also includes an example of a parallel model that does predict additivity.

Proposition 12.3: (i) Mean RT additivity is impossible if

$$\Delta G_{ab}(\tau) \equiv [G_{ab}(\tau; x_{a2}, x_{b2}) - G_{ab}(\tau; x_{a2}, x_{b1})] - [G_{ab}(\tau; x_{a1}, x_{b2}) - G_{ab}(\tau; x_{a1}, x_{b1})] \tag{12.4}$$

is always >0 or always <0 for all $\tau > 0$ and for all $x_{a1} < x_{a2}$ and $x_{b1} < x_{b2}$. Mean RT is sub- or superadditive according to whether $\Delta G_{ab}(\tau) > 0$ or < 0.

(ii) Mean RT additivity is possible if $X_a = X_b = \{0, 1\}$ and if X_a affects S_b as well as S_a, and X_b affects S_a as well as S_b.

Proof: (i) Recall that

$$E[\max(\mathbf{T}_a, \mathbf{T}_b); x_{ai}, x_{bj}] = \int_0^\infty [1 - G_{ab}(\tau; x_{ai}, x_{bj})]\, d\tau, \quad i, j = 1, 2$$

that is, the expectation can be written as the integral of the survivor function.

Therefore, if additivity is to hold for all the x_a and x_b terms, where $x_{i1} < x_{i2}$, $i = a, b$, then

$$\Delta E[\max(\mathbf{T}_a, \mathbf{T}_b)] = E[\max(\mathbf{T}_a, \mathbf{T}_b); x_{a2}, x_{b1}] - E[\max(\mathbf{T}_a, \mathbf{T}_b); x_{a2}, x_{b2}]$$
$$= E[\max(\mathbf{T}_a, \mathbf{T}_b); x_{a1}, x_{b1}] - E[\max(\mathbf{T}_a, \mathbf{T}_b); x_{a1}, x_{b2}]$$

which holds if and only if

$$\int_0^\infty [G_{ab}(\tau; x_{a2}, x_{b2}) - G_{ab}(\tau; x_{a2}, x_{b1})]\, d\tau$$
$$= \int_0^\infty [G_{ab}(\tau; x_{a1}, x_{b2}) - G_{ab}(\tau; x_{a1}, x_{b1})]\, d\tau$$

This equality may be simplified to

$$\int_0^\infty \{[G_{ab}(\tau; x_{a2}, x_{b2}) - G_{ab}(\tau; x_{a2}, x_{b1})] - [G_{ab}(\tau; x_{a1}, x_{b2}) - G_{ab}(\tau; x_{a1}, x_{b1})]\}\, d\tau$$
$$= \int_0^\infty \Delta G_{ab}(\tau)\, d\tau = 0$$

However, by hypothesis (Eq. 12.4), the difference represented by the integrand is always >0 or always <0 so it cannot integrate to 0 and thus additivity is impossible.

The super- or subadditivity follows depending on whether $\Delta G_{ab}(\tau) < 0$ or > 0, because if it is < 0, then, say, the mean RT increases for a given differ-

ence on the X_b dimension will be larger the smaller is $x_a \in X_a$. If $\Delta G_{ab}(\tau) > 0$, then exactly the reverse holds.

(ii) Assume $X_a = X_b = \{0, 1\}$ and consider the following within-stage independent parallel model that is not overall independent. Let $g_{ij}(t; x_a, x_b) =$ parallel density of position $i = a$ or b, during stage $j = 1$ or 2 when x_a and x_b are 0 or 1. Note that we are employing intercompletion time random variables ($\mathbf{T}_i$) on stage i for convenience here rather than total completion time $\mathbf{T}$.

Define

	Position a Density	Position b Density	Condition
Stage 1:	$g_{a1}(t_1; 0, 0) = v \exp(-vt_1)$	$= g_{b1}(t_1; 0, 0)$	$(0, 0)$
Stage 2:	$g_{a2}(t_2; 0, 0) = v \exp(-vt_2)$	$= g_{b2}(t_2; 0, 0)$	
Stage 1:	$g_{a1}(t_1; 0, 1) = v \exp(-vt_1)$	$= g_{b1}(t_1; 0, 1)$	$(0, 1)$
Stage 2:	$g_{a2}(t_2; 0, 1) = 2v \exp(-2vt_2)$	$= g_{b2}(t_2; 0, 1)$	
Stage 1:	$g_{a1}(t_1; 1, 0) = v \exp(-vt_1)$	$= g_{b1}(t_1; 1, 0)$	$(1, 0)$
Stage 2:	$g_{a2}(t_2; 1, 0) = 2v \exp(-2vt_2)$	$= g_{b2}(t_2; 1, 0)$	
Stage 1:	$g_{a1}(t_1; 1, 1) = 2v \exp(-2vt_1)$	$= g_{b1}(t_1; 1, 1)$	$(1, 1)$
Stage 2:	$g_{a2}(t_2; 1, 1) = 4v \exp(-4vt_2)$	$= g_{b2}(t_2; 1, 1)$	

Notice that S_a and S_b are both affected by x_a and x_b in that their densities are speeded up in the first or second stage when x_i $(i = a, b)$ goes from 0 to 1. Letting $\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 = \max(\mathbf{T}_a, \mathbf{T}_b)$ and using methods developed in Chapters 3 and 4, we find that

$$E[\mathbf{T}; (0,0)] - E[\mathbf{T}; (0,1)] = \frac{1}{2v} + \frac{1}{v} - \left(\frac{1}{2v} + \frac{1}{2v}\right) = \frac{1}{2v}$$

$$E[\mathbf{T}; (1,0)] - E[\mathbf{T}; (1,1)] = \frac{1}{2v} + \frac{1}{2v} - \left(\frac{1}{4v} + \frac{1}{4v}\right) = \frac{1}{2v}$$

so additivity indeed holds. □

Proposition 12.4 demonstrates that an important class of independent parallel models form a subset of the class of models addressed in Proposition 12.3(i).

Proposition 12.4: Mean RT additivity is impossible if

$$\begin{aligned} P[\max(\mathbf{T}_a, \mathbf{T}_b) \leq \tau; x_a, x_b] &= P[\mathbf{T}_a \leq \tau \cap \mathbf{T}_b \leq \tau; x_a, x_b] \\ &= G_{ab}(\tau; x_a, x_b) \\ &= G_a(\tau; x_a) G_b(\tau; x_b) \end{aligned}$$

and if

$$\frac{dG_i(\tau; x_i)}{dx_i} > 0, \quad i = a, b \tag{12.5}$$

That is, mean RT additivity is impossible if the total completion time of a and b are independent so that the distribution function on $\max(\mathbf{T}_a, \mathbf{T}_b)$ can be factored, and if G_a and G_b are separate monotonic increasing functions of x_a, x_b, respectively. Thus, the factors X_a and X_b tend to speed up the processing times of a, b, respectively, as they increase.

Proof: Using $\Delta G_{ab}(\tau)$ as developed in Proposition 12.3, we find that the assumption that $G_{ab}(\tau; x_a, x_b)$ is factorable leads to

$$\begin{aligned} \Delta E[\max(\mathbf{T}_a, \mathbf{T}_b)] &= \int_0^\infty \Delta G_{ab}(\tau)\, d\tau \\ &= \int_0^\infty [G_a(\tau; x_{a2}) - G_a(\tau; x_{a1})][G_b(\tau; x_{b2}) - G_b(\tau; x_{b1})]\, d\tau \\ &= 0 \end{aligned}$$

if additivity can occur. Again a contradiction results because by Eq. 12.5 both differences in the product under the integral sign are positive or negative. It is clear that if for both $i = a$ and b

$$\frac{dG_i(\tau; x_i)}{dx_i} > 0$$

then subadditivity follows. □

Thus independence of completion times together with Eq. 12.5 is enough to satisfy the conditions of Proposition 12.3(i) so that the present class of models is an important subset of the Proposition 12.3 class. Observe that Eq. 12.5 implies a strict ordering of distribution functions; e.g., if $x_2 > x_1$, $G_i(\tau; x_2) > G_i(\tau; x_1)$ for all $\tau > 0$.

Actually, the monotonicity for all of X_a and X_b is not really required to produce nonadditivity within a single experiment. As long as there exists a quadruple of values $x_{a1}, x_{a2}, x_{b1}, x_{b2}$ such that $G_a(\tau; x_{a2}) \neq G_a(\tau; x_{a1})$ and $G_b(\tau; x_{b2}) \neq G_b(\tau; x_{b1})$ for all $\tau \geq 0$, that experiment will yield nonadditivity.

Note further that Propositions 12.3 and 12.4 form a sort of "dual" to Proposition 12.1. Part (i) of Proposition 12.3 and Proposition 12.4 say that if the two subsystems operate in such a way that X_a and X_b separately and monotonically affect their respective subsystem distribution functions, then mean additivity is impossible. Part (ii) of Proposition 12.3 states that if X_a and/or X_b affect both subsystems, then mean RT additivity *may* occur. In particular, it is important to observe that the effects tend to act in reverse: Whereas separate action of X_a and X_b tends to promote additivity in serial systems, it tends to promote interactions in parallel systems. Conversely, overlapping action can produce interactions in serial systems and can produce additivity in parallel systems.

Furthermore, an instance of mean RT additivity in *independent* parallel models arises in a parallel analog to the standard exponential serial model for memory scanning outlined in a formal way earlier in this chapter. The usual

parallel concomitant to the standard serial model produces *dependent* completion times (this is the reallocation parallel model). To start, just consider two serial positions a and b and let $X_i=\{0,1\}$, $i=a,b$, where 0 refers to the absence of an item in location a or b and 1 refers to the presence of an item. Suppose that when $x_i=0$, the contribution of S_i $(i=a,b)$ to RT is 0. In condition $(1,0)$ and $(0,1)$ assume the S_a and S_b processing time densities are each ve^{-vt}. In condition $(1,1)$ assume both the S_a and S_b densities are $\frac{3}{4}v\exp(-\frac{3}{4}vt)$. The mean RT predictions under the assumed parallel independent processing are then

Condition	$\overline{RT}$
$(0,0)$	0
$(0,1)$	$1/v$
$(1,0)$	$1/v$
$(1,1)$	$2/v$

just as in the standard serial model in the special Donderian case of the additive factor method. For instance, for $(0,1)$, the contribution of S_a is 0 and S_b is $1/v$. In $(1,1)$ the expected exhaustive time is (see Chapter 4 for detail)

$$E[\max(\mathbf{T}_a,\mathbf{T}_b)]=\frac{1}{v_a}+\frac{1}{v_b}-\frac{1}{v_a+v_b}$$
$$=\frac{4}{3v}+\frac{4}{3v}-\frac{1}{3v/4+3v/4}=\frac{8}{3v}-\frac{4}{6v}$$
$$=\frac{6}{3v}=\frac{2}{v}$$

Note that X_a affects S_b as well as S_a and similarly for X_b, in the sense that when $x_a=1$ the value of the S_a rate parameter depends on X_b. Thus, when $x_a=1$, $x_b=0$, $v_a=v$, but when $x_a=1$, $x_b=1$, then $v_a=\frac{3}{4}v$, so that the assumption of selective influence of the experimental factors is violated. It is shown in Chapter 4 that for n items in memory, the present independent parallel model is extended by letting the exponential rate be

$$v(n)=\frac{u}{n}\sum_{i=1}^{n}\frac{1}{i}$$

where u is the serial rate.

Finally, observe in the proof of part (i) of Proposition 12.3 and Proposition 12.4 that mean RT additivity is impossible even for a single set of values $\{x_a,x_b,\Delta x_a,\Delta x_b\}$ so that any single experiment should falsify additivity within the treated type of models, barring, of course, problems of a purely experimental or statistical nature.

What characteristics should independent parallel models have that might be able to predict additivity in the event that the condition is dropped that G_a and G_b are monotonic functions of X_a and X_b, respectively? The assumption that X_i affects only S_i $(i=a,b)$ continues. The remaining part of this subsection will discuss this question. The results to date are somewhat heuristic, so the reader may wish to skip ahead to the next subsection on a first reading.

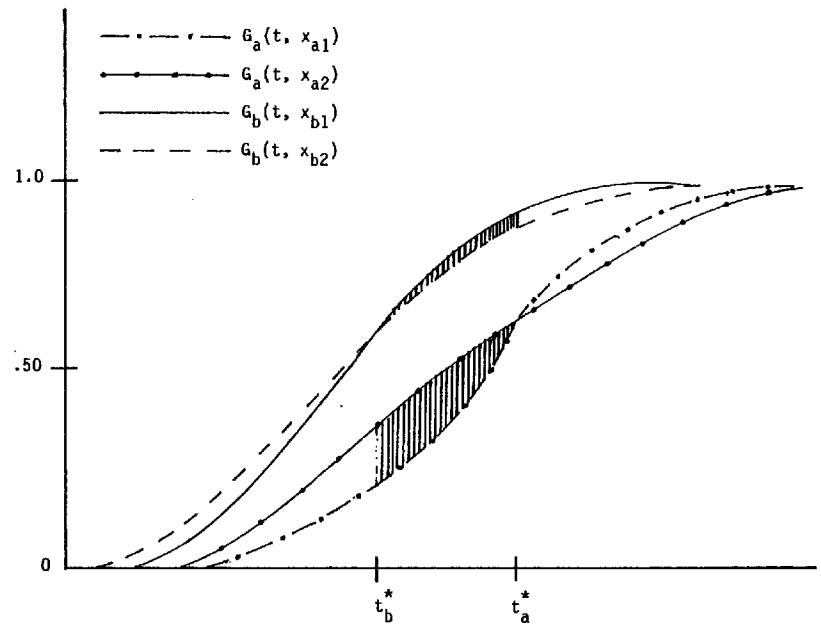


Fig. 12.5. Independent parallel distributions indicating how distributions that are not monotonic functions of their respective factors might predict mean RT additivity. (Curves and arcs are approximate and for illustration only.)

First note that under the second part of Eq. 12.5

$$G_i(\tau;x_i)<G_i(\tau;x_i') \quad \text{if and only if } x_i<x_i' \text{ for all } 0<\tau; \quad i=a,b$$

If the monotonicity assumption and therefore the above inequality are violated for a specific set $x_{a1},x_{a2},x_{b1},x_{b2}$, an implication is that there exist values of τ for which, say, $G_a(\tau;x_{a1})<G_a(\tau;x_{a2})$ but others for which $G_a(\tau;x_{a1})\geqslant G_a(\tau;x_{a2})$ for $x_{a1}\neq x_{a2}$ (where x_{a1} and $x_{a2}\in X_a$). What kind of crossovers of the parallel distribution functions might predict a lack of interactions in the data? Let us analyze the constraints imposed by Proposition 12.3, under the assumption that only one crossover occurs. Figure 12.5 will help the reader to interpret the analysis. It is also assumed that $G_i(\tau;x_{i1})<G_i(\tau;x_{i2})$ for either subsystem before the crossover but $G_i(\tau;x_{i1})>G_i(\tau;x_{i2})$ $(i=a,b)$ after the crossover. The corresponding intuition here is that the mean associated with the X_{i1} factor will be less than that associated with the X_{i2} factor, because even though the X_{i1} distribution is climbing slower during early τ-values, the X_{i2} distribution has a long, higher tail and thus produces a

larger mean processing time. Of course, it need not be that the crossover point is the same for both S_a and S_b.

We shall analyze the case where the crossover point for S_b is earlier than that for S_a, but the other cases may be handled similarly. Call the crossover point for the a curves τ_a^* and that for the b curves τ_b^*. Then we can describe the constraining equality involved in the proof of Proposition 12.3(i) as

$$\int_0^\infty G_a(\tau;x_{a1})[G_b(\tau;x_{b1})-G_b(\tau;x_{b2})]\,d\tau$$
$$=\int_0^\infty G_a(\tau;x_{a2})[G_b(\tau;x_{b1})-G_b(\tau;x_{b2})]\,d\tau$$

Let $G_a(\tau;x_{ai})=A_i(\tau)$ and $G_b(\tau;x_{bi})=B_i(\tau)$, and then a little manipulation permits the equality to be expressed as

$$\int_0^{\tau_b^*}[B_1(\tau)-B_2(\tau)][A_1(\tau)-A_2(\tau)]\,d\tau+\int_{\tau_a^*}^\infty[B_1(\tau)-B_2(\tau)][A_1(\tau)-A_2(\tau)]\,d\tau$$
$$=\int_{\tau_b^*}^{\tau_a^*}[B_2(\tau)-B_1(\tau)][A_1(\tau)-A_2(\tau)]\,d\tau$$

All three integrals here are positive. The interpretation is that the integral from τ_b^* to τ_a^* has to compensate for the fact that from τ_a^* to ∞ the positive A-difference is weighted *more* by B_1 than by B_2, and from 0 to τ_b^* the negative A-difference is weighted *less* by B_1 than by B_2, so that the middle region (the shaded areas in Fig. 12.5) has to counteract this tendency to produce an inequality. There seems to be little doubt that distribution functions satisfying this condition exist, because within the constraint of monotonic increasing functions, one is free to manipulate the form of the functions. However, probability functions are not specified in such a way as to make the generation of such distributions trivial. Too, the question is largely academic, because the a priori likelihood that distributions with this property (i.e., having just exactly the proper crossover points) would be found in natural systems is very probably vanishingly small.

A characteristic function approach to additivity in parallel models

An alternative approach to the two-factor manipulation gives additivity not only of the means but of all cumulants. It will be recalled that if $K(s)$ is the cumulant function, then $\exp[K(s)]$ equals the characteristic function. This approach will therefore be much more confining than when only mean RT additivity is required. We begin with a single "experiment" under factor levels (x_{ai}, x_{bj}), where $i,j=1,2$.

Let $C_{a1b1}(s)$ be the characteristic function (hereafter referred to as a C function) of the joint parallel distribution on exhaustive processing times under specific experimental factor levels x_{a1} and x_{b1}. Similarly, let $C_{a2b1}(s)$ and $C_{a1b2}(s)$ be the C functions obtained by changing the factors X_a and X_b to x_{a2} and x_{b2}, respectively, while leaving the others unaffected.

Note that $C_{a2b2}(s)$ is the C function of the joint distribution resulting from changing X_a and X_b to x_{a2} and x_{b2} at the same time. These correspond to the four parts of Eq. 12.6 below. Finally, let

$$K(s)=\sum_{i=1}^{\infty} k_i(s)\frac{s^i}{i!}$$

be a cumulant function; then $k_i(s)$ is said to be the ith cumulant of the distribution represented by $K(s)$.

The outline of the approach is that $C_{a2b1}(s)$ and $C_{a1b2}(s)$ can always be expressed in terms of $C_{a1b1}(s)$ and another function $C_j'(s)$ ($j=a$ or b) even when the latter is not a true C function. There is no bite at all in these formulations so far. However, we are able to characterize the demand that the effects of x_{a2} and x_{b2}, when applied together, work additively in terms of their effects perpetrated separately (e.g., as does x_{a2} when paired with x_{b2}). These equations can be written:

Characteristic functions	*Level of factors*		
$C_{a1b1}(s)$	x_{a1}, x_{b1}	(i)	
$C_{a2b1}(s)=C_{a1b1}(s)\,C_a'(s)$	x_{a2}, x_{b1}	(ii)	
$C_{a1b2}(s)=C_{a1b1}(s)\,C_b'(s)$	x_{a1}, x_{b2}	(iii)	
$C_{a2b2}(s)=C_{a1b1}(s)\,C_a'(s)\,C_b'(s)$	x_{a2}, x_{b2}	(iv)	(12.6)

Proposition 12.5: A parallel model predicts additivity of all cumulants whenever

$$C_{a2b2}(s)=\frac{C_{a2b1}(s)\,C_{a1b2}(s)}{C_{a1b1}(s)} \tag{12.7}$$

Proof: Simply solve (ii) and (iii) in Eq. 12.6 for $C_a'(s)$ and $C_b'(s)$, respectively, and substitute these into (iv) to achieve the desired result, Eq. 12.7.

Now, a cumulant function under standard regularity conditions may be written in series form as

$$K(s)=\sum_{i=1}^{\infty} k_i(s)\frac{s^i}{i!}$$

so that

$$C(s)=e^{K(s)}=\exp\left[\sum_{i=1}^{\infty} k_i(s)\frac{s^i}{i!}\right]$$

Therefore, dividing both sides of Eq. 12.7 by $C_{a2b1}(s)$ yields

$$\frac{C_{a2b2}(s)}{C_{a2b1}(s)}=\exp[K_{a2b2}(s)-K_{a2b1}(s)]\equiv\exp[K_{a1b2}(s)-K_{a1b1}(s)]$$

that is,

$$\exp\left\{\sum_{i=1}^{\infty}\left[k_{a2b2i}(s)-k_{a2b1i}(s)\right]\frac{s^i}{i!}\right\}\equiv\exp\left\{\sum_{i=1}^{\infty}\left[k_{a1b2i}(s)-k_{a1b1i}(s)\right]\frac{s^i}{i!}\right\}$$

It then follows by standard arguments that

$$k_{a2b2i}(s)-k_{a2b1i}(s)=k_{a1b2i}(s)-k_{a1b1i}(s),\quad i=1,2,\ldots,$$

and for all s in the domain of k. This establishes additivity in the cumulants. □

Naturally, because the mean and variance *are* the first two cumulants k_1 and k_2, these are also additive. In order to come up with the density function corresponding to $C_{a2b2}(s)$ [call it $g_{a2b2}(t)$], one must take the inverse Fourier transform of that function. Even when this is feasible, the resulting density may have no immediately transparent parallel interpretation and certainly may not be factorable into two separate parallel densities as would be required if independence were to hold. It is even less likely that a particular chosen parallel interpretation for $g_{a2b2}(t)$ will, for example, possess parameter structure such that those parameters associated with the $a2$ part of the distribution are only functions of X_a and the $b2$ part only of X_b.

We do have an example that yields an independent and therefore factorable $G_{a2b2}(t)=G_{a2}(t)\cdot G_{b2}(t)$, but without satisfying the last requirement mentioned above. It must be emphasized that this example is *only* for a single "experiment" in the sense that it works only for a particular set of values $x_{a1}, x_{b1}, x_{a2}, x_{b2}$. Readers wishing to move quickly to results concerned with satisfaction of Eq. 12.7 for all $(x_a, x_b)\in X_a\times X_b$ may skip to the following subsection.

Assume that the distributions of (i), (ii), and (iii) of Eq. 12.6 are exponential and independent and that $v_a(x_{ai})=x_{ai}$ and $v_b(x_{bi})=x_{bi}$ are the respective rate parameters for these first three distributions. Let $x_{a1}=x_{b1}=8$, $x_{a2}=4$ and $x_{b2}=2$. Then the pertinent distribution functions are:

$$G_{a1b1}(t)=(1-e^{-8t})(1-e^{-8t})=G_{a1}(t)G_{b1}(t)$$
$$G_{a2b1}(t)=(1-e^{-4t})(1-e^{-8t})=G_{a2}(t)G_{b1}(t)$$
$$G_{a1b2}(t)=(1-e^{-8t})(1-e^{-2t})=G_{a1}(t)G_{b2}(t)$$

Next we calculate the associated characteristic functions:

$$C_{a1b1}(s)=\frac{16}{8+is}-\frac{16}{16+is}$$
$$C_{a2b1}(s)=\frac{4}{4+is}+\frac{8}{8+is}-\frac{12}{12+is}$$
$$C_{a1b2}(s)=\frac{8}{8+is}+\frac{2}{2+is}-\frac{10}{10+is}$$

and then use Proposition 12.5 to formulate $C_{a2b2}(s)$:

$$C_{a2b2}(s)=\frac{C_{a2b1}(s)C_{a1b2}(s)}{C_{a1b1}(s)}=\frac{[4/(4+is)+8/(8+is)-12/(12+is)]\times[8/(8+is)+2/(2+is)-10/(10+is)]}{16/(8+is)-16/(16+is)}$$

The inverse Fourier transform of this C function is $g_{a2b2}(t)=13.2e^{-12t}+3.8e^{-2t}+24e^{-8t}-6e^{-4t}-3.5e^{-10t}$ and the distribution function is $G_{a2b2}(t)=1-1.1e^{-12t}-1.9e^{-2t}-3e^{-8t}+1.5e^{-4t}+3.5e^{-10t}$, and it turns out that this can be factored into

$$G_{a2b2}(t)=G_{b2}(t)G_{a2}(t)$$
$$=(1-e^{-2t})(1-.9e^{-2t}+.6e^{-4t}+.6e^{-6t}-2.4e^{-8t}+1.1e^{-10t})$$

Hence, in the present instance, not only is $G_{a2b2}(t)$ an acceptable distribution function, but it is composed of two independent parallel distribution functions. Nevertheless, as anticipated, this particular characterization does not meet one of the standards we have been requiring, namely that $G_{a2}(t)$ should not be a function of X_b, although $G_{b2}(t)$ *is* correctly equal to $1-e^{-2t}$ (a function only of $X_{b2}=2$). The intuition, certainly, of this state of affairs, in terms of the underlying system, would be that the factors are inappropriately influencing the wrong subsystems. It is possible that another factorization exists that does not possess this failing, but we have not been able to find one.

Characteristic functions for a continuum of experiments

Suppose now that rather than considering the condition $C_{a2b2}(s)$ as pertinent only for two distinct values of each of $x_a\in X_a$ and $x_b\in X_b$, as done implicitly above, we assume that it holds for all x_a, x_b and all increments (or decrements less than or equal to x_a or x_b), and we continue to assume that $X_a=\mathbf{R}^+$ and $X_b=\mathbf{R}^+$ as well. Thus, for any experiment we run, additivity must hold. This suggests we write

$$C_{a2b2}(s)=C(s,x_a+h_a,x_b+h_b),\quad C_{a2b1}(s)=C(s,x_a+h_a,x_b),$$
$$C_{a1b2}(s)=C(s,x_a,x_b+h_b),\quad\text{and}\quad C_{a1b1}(s)=C(s,x_a,x_b)$$

Now the germane additivity condition of Proposition 12.5 becomes

$$C(s,x_a+h_a,x_b+h_b)=\frac{C(s,x_a+h_a,x_b)C(s,x_a,x_b+h_b)}{C(s,x_a,x_b)}$$

or

$$C(s,x_a+h_a,x_b+h_b)C(s,x_a,x_b)=C(s,x_a+h_a,x_b)C(s,x_a,x_b+h_b)\qquad(12.8)$$

We are now in a position to offer Proposition 12.6, which shows that satisfaction of Eq. 12.8 implies that the basic C function $C(s,x_a,x_b)$ must be factorable into separate functions of x_a and x_b, under certain analytic conditions.

Proposition 12.6: If Eq. 12.8 holds and if the first and second partial derivatives of $Q=\ln(C)$ exist and are bounded and continuous on X_a and X_b, then

$$C(s, x_a, x_b) = C_a(s, x_a)\, C_b(s, x_b)$$

Thus, if C_a and C_b are true C functions, then $C(s, x_a, x_b)$ is equivalent to a serial stochastically independent model.

Proof: Set $Q = \ln(C)$ and Eq. 12.8 becomes

$$Q(s, x_a + h_a, x_b + h_b) + Q(s, x_a, x_b) = Q(s, x_a + h_a, x_b) + Q(s, x_a, x_b + h_b)$$

Subtracting the right side from the left, dividing by $h_a h_b$ and taking the limits produces

$$0 = \lim_{h_b \to 0} \left\{ \frac{\lim\limits_{h_a \to 0} \dfrac{\begin{array}{c}[Q(s, x_a + h_a, x_b + h_b) - Q(s, x_a + h_a, x_b)] \\ - [Q(s, x_a, x_b + h_b) - Q(s, x_a, x_b)]\end{array}}{h_a}}{h_b} \right\}$$

$$= \lim_{h_b \to 0} \left\{ \frac{\dfrac{\partial Q(s, x_a, x_b + h_b)}{\partial x_a} - \dfrac{\partial Q(s, x_a, x_b)}{\partial x_a}}{h_b} \right\}$$

$$= \frac{\partial^2 Q(s, x_a, x_b)}{\partial x_a \partial x_b} \quad \text{for all } (x_a, x_b) \in \mathbf{R}^+ \times \mathbf{R}^+$$

Therefore, Q must be linear in X_a and X_b:

$$Q(s, x_a, x_b) = \alpha f(s) + \beta g(s, x_a) + \gamma h(s, x_b) = \beta g'(s, x_a) + \gamma h'(s, x_b)$$

where

$$g'(s, x_a) = g(s, x_a) + \frac{\alpha}{\beta + \gamma} f(s) \quad \text{and} \quad h'(s, x_b) = h(s, x_b) + \frac{\alpha}{\beta + \gamma} f(s)$$

Returning to the original C functions, the linearity in $\ln(C)$ corresponds to multiplicativity in C:

$$C(s, x_a, x_b) = C_a(s, x_a)\, C_b(s, x_b)$$

where

$$C_a(s, x_a) = \exp[\beta g'(s, x_a)] \quad \text{and} \quad C_b(s, x_b) = \exp[\gamma h'(s, x_b)]$$

If C_a and C_b are true C functions, the inverse Fourier transform of $C(s, x_a, x_b)$ will now be a convolution of two independent densities, thereby defining a serial stochastically independent model. □

This outcome is very close to confining the class of acceptable models to the serial class because we end with a product of separate functions of X_a, X_b, which of course defines a convolution of two separate temporal functions of X_a and X_b in the time domain. The only shortcoming is that we do not know for sure that all the C_a and C_b are true characteristic functions; so

although the convolution in the time domain will itself overall be a well-defined distribution, the separate time functions may, in principle, not be.

An important question, therefore, is whether there are interesting parallel models that obey this stricture. We doubt that there are, especially considering the difficulties we encountered in constructing parallel examples that simply satisfied Proposition 12.5 for a *single set* $(x_{a1}, x_{b1}, x_{a2}, x_{b2})$ or were capable of yielding even mean RT additivity. However, these questions deserve more work.

Examples of additive serial systems and "falsified" parallel models

Suppose S_a is governed by densities $f_{a1}(t_{a1}; x_a)$ and $f_{a2}(t_{a2}; x_a)$ for stages 1 and 2, and S_b is governed by densities $f_{b1}(t_{b1}; x_b)$ and $f_{b2}(t_{b2}; x_b)$, and that the system is serial with subsystem S_a doing its job first with probability p. The p parameter has been equal to 1 above, but the added generality will cause no difficulty here. Across-stage independence (here equivalent to stochastic independence) is assumed.

Across-stage independence is also equivalent to assuming that all the intercompletion times are independent. The joint density function on intercompletion times is, of course,

$$f_{ab}(t_a, t_b; x_a, x_b) = p f_{a1}(t_{a1}; x_a) f_{b2}(t_{b2}; x_b) + (1-p) f_{b1}(t_{b1}; x_b) f_{a2}(t_{a2}; x_a)$$

and the exhaustive completion time ($\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2$) distribution function is

$$P_{ab}(\mathbf{T} \leqslant t) = p \int_0^t \int_0^{t - t_{a1}} f_{a1}(t_{a1}; x_a) f_{b2}(t_{b2}; x_b)\, dt_{b2}\, dt_{a1} + (1-p) \int_0^t \int_0^{t - t_{b1}} f_{b1}(t_{b1}; x_b) f_{a2}(t_{a2}; x_a)\, dt_{a2}\, dt_{b1}$$

$$= p \int_0^t f_{a1}(t_{a1}; x_a) F_{b2}(t - t_{a1}; x_b)\, dt_{a1} + (1-p) \int_0^t f_{b1}(t_{b1}; x_a) F_{a2}(t - t_{b1}; x_a)\, dt_{b1}$$

The average mean RT (as usual ignoring other residual processing durations),

$$E_{ab}(\mathbf{T}_a + \mathbf{T}_b; x_a, x_b) = p[E_a(\mathbf{T}_{a1}; x_a) + E_b(\mathbf{T}_{b2}; x_b)] + (1-p)[E_b(\mathbf{T}_{b1}; x_b) + E_a(\mathbf{T}_{a2}; x_a)]$$

is naturally an additive function of the X_a and X_b scales.

Let us next focus on a general but fairly tractable parallel model, which has the property of within-stage independence but that is not necessarily exponential. Because the serial class we shall be trying to mimic has complete

across-stage independence (i.e., independence of intercompletion times), we will also assume this holds for the parallel class. This allows us to investigate equivalence of each stage (each intercompletion time distribution) separately. In fact, such equivalence is necessary and sufficient for overall distribution equivalence. Furthermore, we saw in Chapter 4 that when $n=2$, that is, for two elements or two subsystems, all the onus falls on stage 1 when both elements are uncompleted, because at stage 2 the parallel and serial distributions can be made equivalent by fiat.

Therefore, we first look at the first stage, or minimum time distribution. We saw in Chapter 4 (Prop. 4.4) that when it exists, the solution of the parallel distribution in terms of the serial is of the form (see also Townsend 1976b)

$$\bar{G}_a(t;x_a)=\exp\left[-\int_0^t \frac{pf_a(t';x_a)\,dt'}{p\bar{F}_a(t';x_a)+(1-p)\bar{F}_b(t';x_b)}\right]$$
$$\bar{G}_b(t;x_b)=\exp\left[-\int_0^t \frac{(1-p)f_b(t';x_b)\,dt'}{p\bar{F}_a(t';x_a)+(1-p)\bar{F}_b(t';x_b)}\right] \tag{12.9}$$

where $\bar{G}_i(t;x_i)$ is the parallel survivor function $[1-G_i(t;x_i)]$ for subsystem S_i $(i=a,b)$, $\bar{F}_i(t;x_i)$ is the concomitant serial survivor function $[1-F_i(t;x_i)]$, and $f_i(t;x_i)$ $(i=a,b)$ is, of course, the serial density function. Note that we have dropped the stage designation for convenience because we are now concentrating on stage 1. Now, $\bar{G}_a$ and $\bar{G}_b$, when the solution to Eq. 12.9 exists, are each functions of *both* x_a and x_b, violating the earlier stricture against the x_a affecting only the S_a distribution, and similarly for x_b and S_b. Further, there are cases where $\bar{G}_a$ and/or $\bar{G}_b$ cannot exist when p, f_a, and f_b do. Thus, at the distributional level, it is possible in principle to falsify all within-stage independent parallel models.

However, let us concentrate on the class of independent parallel models, which are, of course, special cases of the class of within-stage independent parallel models, and consider the exhaustive completion time.

Here it will be convenient to use $\mathbf{T}$ for both the parallel and serial models. In this case, employing $\mathbf{T}=\mathbf{T}_1+\mathbf{T}_2=$exhaustive completion time,

$$P_{ab}(\mathbf{T}\leqslant t)=G_a(t;x_a)G_b(t;x_b)=G_{ab}(t;x_a,x_b)$$

Interestingly, $\log P_{ab}(\mathbf{T}\leqslant t)$ is therefore additive in x_a and x_b for any $t\geqslant 0$. Thus, the log distribution function could be checked for this additivity, which is *not* predicted by the serial model. Nevertheless, this prediction would be somewhat nullified by the necessary convolution of the duration of residual subsystems outside S_a and S_b. It would be preferable to factor out such processes, if possible, before such a test. As we learned earlier in the chapter, independent parallel models also cannot predict mean RT additivity under reasonably broad conditions.

We now narrow the focus to models based on exponential intercompletion times and also obeying the above restrictions, namely that the serial models assume stochastic independence and the parallel models assume independence of total completion times.

Set

$$p=1,$$
$$f_a(t_a;x_a)=u_a(x_a)\exp[-u_a(x_a)t_a],$$
$$f_b(t_b;x_b)=u_b(x_b)\exp[-u_b(x_b)t_b],$$
$$\text{for } x_a\in X_a=\mathbf{R}^+,\quad x_b\in X_b=\mathbf{R}^+ \tag{12.10}$$

We know from earlier chapters that parallel models, even nonindependent ones, cannot be equivalent to serial models when $p=1$, although they may (depending on u_a, u_b) approximate them by beginning to turn into a serial model. However, even when $p\neq 0$ or 1, it is known that if $u_a(x_a)\neq u_b(x_b)$, no parallel model based on within-stage independence is equivalent to the serial model expressed by Eq. 12.10 (Townsend 1976b; see also Chapter 4); that is, not both of $\bar{G}_a$ and $\bar{G}_b$ can exist. But the ability to separately manipulate $u_a(x_a)$ and $u_b(x_b)$ is the fundamental assumption of the additive factor method, and therefore the present independent parallel models cannot completely mimic (be equivalent in distribution to) the serial model, which is producing additivity in x_a and x_b.

For simplicity of notation, let the dependence on x_a and x_b be implicit, so we write $u_i(x_i)=u_i$ $(i=a,b)$ in the serial and $v_i(x_i)=v_i$ $(i=a,b)$ in the parallel models.

The true exhaustive serial density is, then,

$$f_{ab}(t;u_a,u_b)=p\int_0^t f_{a1}(t_{a1})f_{b2}(t-t_{a1}\,|\,t_{a1})\,dt_{a1}$$
$$+(1-p)\int_0^t f_{b1}(t_{b1})f_{a2}(t-t_{b1}\,|\,t_{b1})\,dt_{b1}$$
$$=\frac{u_a\cdot u_b}{u_a-u_b}[\exp(-u_b t)-\exp(-u_a t)]$$

The ensuant distribution function is

$$F_{ab}(t;u_a,u_b)=1-\frac{1}{u_a-u_b}[u_a\exp(-u_b t)-u_b\exp(-u_a t)]$$

On the other hand, the exhaustive independent parallel model density with exponential times is just

$$g_{ab}(t;v_a,v_b)=\int_0^t v_a\exp[-(v_a+v_b)t_{a1}]v_b\exp[-v_b(t-t_{a1})]\,dt_{a1}$$
$$+\int_0^t v_b\exp[-(v_a+v_b)t_{b1}]v_a\exp[-v_a(t-t_{b1})]\,dt_{b1}$$
$$=v_b\exp(-v_b t)[1-\exp(-v_a t)]+v_a\exp(-v_a t)[1-\exp(-v_b t)]$$
$$=g_b(t)G_a(t)+g_a(t)G_b(t)$$

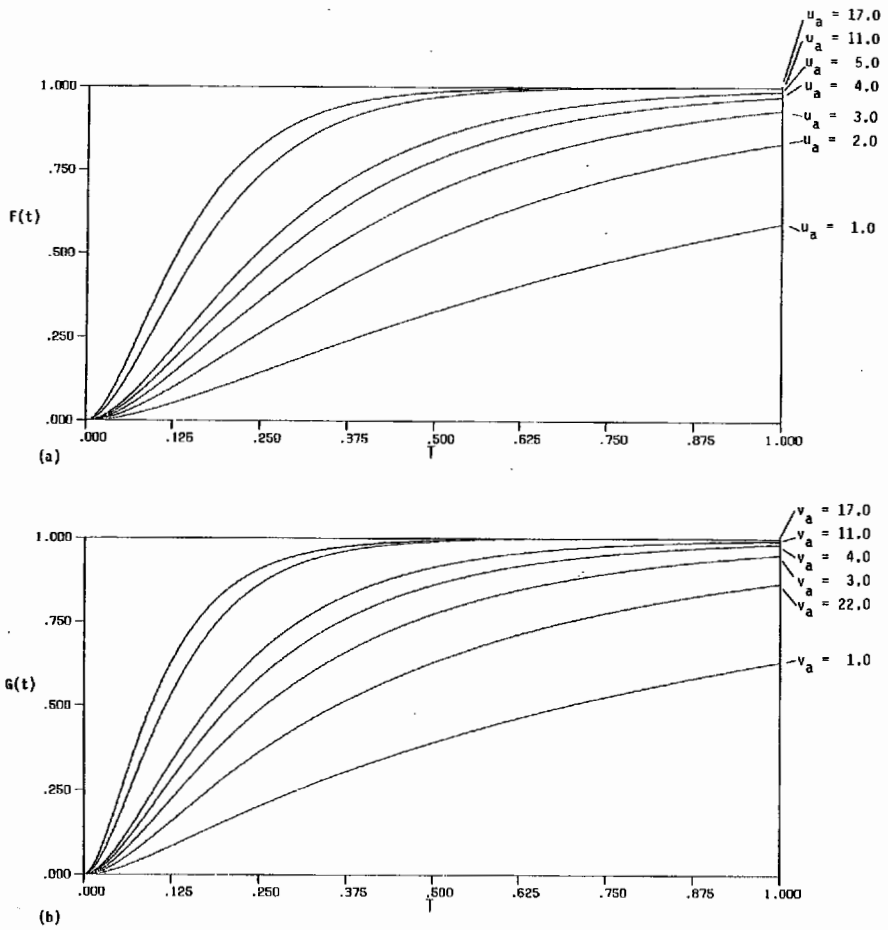


Fig. 12.6. (a) Serial exhaustive distribution functions as u_a is manipulated and u_b is held constant. The serial distribution function used is

$$F(t)=1-\frac{u_a\exp(-u_b t)-u_b\exp(-u_a t)}{(u_a-u_b)}, \quad \text{with } u_b=10.00.$$

(b) Parallel exhaustive distribution functions as v_a is manipulated and v_b is held constant. The parallel distribution function used is

$$G(t)=[1-\exp(-v_a t)][1-\exp(-v_b t)], \quad \text{with } v_b=10.0.$$

and the distribution function is, of course,

$$G_{ab}(t; v_a, v_b)=G_a(t; v_a)\,G_b(t; v_b)=[1-\exp(-v_a t)][1-\exp(-v_b t)]$$

Figure 12.6a shows how the serial distribution functions change as $u_a(x_a)$ is varied for a fixed $u_b(x_b)=10$ and $p=1$, and Fig. 12.6b does the same for the present parallel model, with $v_b(x_b)=10$ and varying $v_a(x_a)$.

Although we are in the presence of two distributions that cannot be equivalent, and although we have arbitrarily let $p=1$ without attempting to use it to make the exhaustive time distributions look more alike, the graphs of $G_{ab}(t)$ and $F_{ab}(t)$ appear rather similar. Recall, however, that $\log G(t)$ is additive in v_a and v_b. Therefore, let us plot

$$\log G_{ab}(t; v_a=10, v_b=10)-\log G_{ab}(t; v_a=1, v_b=10)$$

and compare it with the plot of

$$\log G_{ab}(t; v_a=1, v_b=10)-\log G_{ab}(t; v_a=1, v_b=1)$$

and do the same for the serial distributions. The results are shown in Fig. 12.7. Note that the two curves coincide in the case of the parallel functions but are different in the case of the serial functions. Alternatively, one can hold t constant and plot $\log G_{ab}(t; v_a, v_b)$ for varying values of v_a and v_b: The result will be parallel functions here, but not for the serial model.

Thus, in principle at least, independent parallel distributions on exhaustive processing time can be tested against serial exhaustive distributions by examining additivity of the logarithm of the distribution function. However, in the absence of a method to get rid of the residual (leftover) RT process distribution, this will work only if the residual times contribute little variance to the overall RT distribution (see also Ashby & Townsend 1980 for work on a similar problem).

Finally, the mean RT predictions in the serial case are

$$E_{ab}(\mathbf{T}; u_a, u_b)=\frac{1}{u_a}+\frac{1}{u_b}$$

transparently additive in u_a and u_b. On the other hand, in the parallel model we have

$$E_{ab}(\mathbf{T}; v_a, v_b)=\frac{1}{v_a}+\frac{1}{v_b}-\frac{1}{v_a+v_b}$$

which is clearly an interactive prediction as a function of v_a and v_b.

Reaction time and measurement theory

Although there has been considerable work done applying principles of modern measurement theory to psychophysics, there is still relatively little interplay between it and research in information processing and cognitive psychology. It seems likely that such interactions will increase over the next few years.

Perhaps the first problem that must be solved before measurement theory can be usefully applied to the study of RT, is to determine what kind of scale RT lies on. Many, like Sternberg (1969a, b), believe that only the physical scale (a ratio scale, presumably) of time should be used. Others, like Micko (1969), are of the opinion that RT distributions relate to a psychological time scale that differs from the physical time scale and that a psychological func-

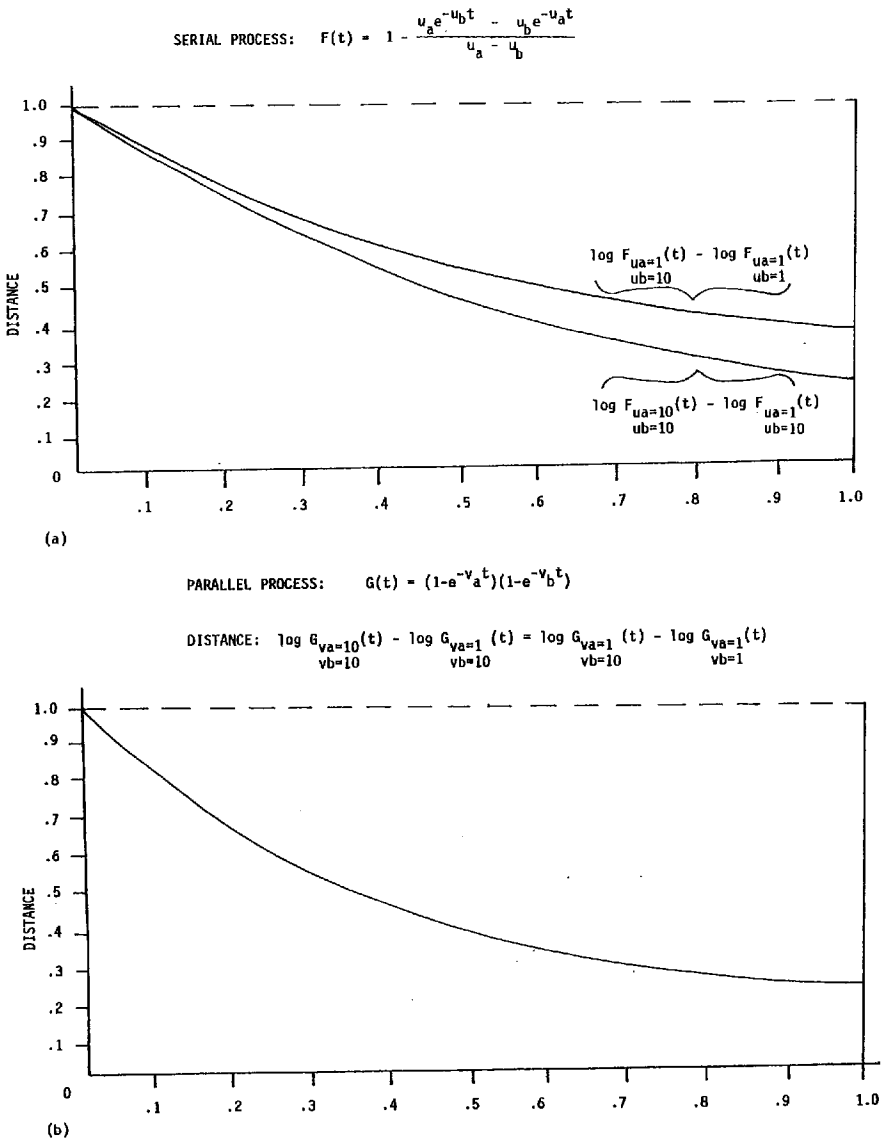


Fig. 12.7. (a) Logarithms of the serial exhaustive distribution functions for two values of u_a and u_b. (b) Logarithms of the parallel exhaustive distribution functions for two values of v_a and v_b.

tion can be derived that connects that physical time scale to the psychological time scale. Still others appear to feel that in most experiments, RT should be treated as lying on, at best, an ordinal scale (if so, it is problematic whether it is valid to compute and compare mean RTs).

An answer to this problem requires a little background in the fundamentals of modern measurement theory. Consider some empirical system E with a set of real world objects A, an empirical relation $\succsim$ defined on the elements of A and perhaps some empirical operation o, such as concatenation. In an example from extensive measurement, A might be the set of all rods, with $a_1 \succsim a_2$ if and only if, when the rods a_1 and a_2 are laid side by side, rod a_1 protrudes at least as far as rod a_2 (i.e., a_1 is at least as long as a_2). The empirical operation o is interpreted as concatenation so that $a_1 o a_2$ means that the rods a_1 and a_2 are laid end to end.

The representation problem of measurement theory is to find a numerical system N consisting of a numerical set R, a relation $\geqslant$, and an operation (say, $+$) such that the triple $N = \langle R, \geqslant, + \rangle$ is homomorphic to the empirical system $E = \langle A, \succsim, o \rangle$, that is, so that it mimics the properties of E that we wish to measure. In the rod example, a homomorphic numerical system is $\langle \mathbf{R}^+, \geqslant, + \rangle$, where $\mathbf{R}^+$ is the set of nonnegative real numbers.

Of course, to establish such a homomorphism one must exhibit a function f that maps the empirical elements of A into the numerical elements of R. Thus f assigns a number to each element of A. The homomorphism guarantees that the numbers assigned to the objects are related to each other through $\geqslant$ and $+$ in the same way that the elements of A are related through $\succsim$ and o. The triple $\langle E, N, f \rangle$ is called a *measurement scale.*

Suppose now that the representation problem has been solved, so that some homomorphic numerical system has been found. Typically, more than one such homomorphic system exists. Thus if f is a homomorphic mapping from E to N, there will usually exist many such homomorphic mappings. The uniqueness problem is to find *all* numerical systems that are homomorphic to the given empirical system and their relation to one another. Usually, this is accomplished by specifying properties of the transformation ϕ that maps one such numerical system into another. For example, an interval scale of measurement has the property that ϕ is linear. Thus if $f: E \to N$ is a homomorphic mapping and $\langle E, N, f \rangle$ is an interval scale, then $\phi(f) = af + b: E \to N$, where a and b are constants, is another homomorphic mapping. For a more detailed introduction to the fundamentals of measurement theory the reader is referred to Suppes and Zinnes (1963), Krantz, Luce, Suppes, and Tversky (1971), or Roberts (1979).

In the case of RT, the empirical relation $\succsim$ can assumedly be interpreted as something like "requires at least as much processing time as." Unfortunately, this ordering is not directly observable. If it were, it would be an easy matter to determine what scale RT, or more precisely, processing time, lies on. One would merely have to check whether the empirical processing time system satisfies the axioms of ordinal, interval, or ratio scale measurement.

Instead, it is typically assumed that a task X_a requires at least as much processing time as a task X_b if and only if $\mathrm{RT}(X_a) \geqslant \mathrm{RT}(X_b)$, or at least if some probabilistic counterpart of this RT ordering holds. In addition, it is widely accepted that addition is a meaningful operation to perform on RTs, and

thus the great majority of RT theorists implicitly assume that in the case of processing time, the relevant homomorphic numerical system is $\langle RT, \geqslant, + \rangle$, where RT is the set of nonnegative reals.

It can easily be shown that this numerical structure satisfies all the axioms of extensive measurement and thus by recording RT and by using the relevant field properties associated with the structure $\langle RT, \geqslant, + \rangle$, we are assuming that RT lies on a ratio scale. As long as the empirical system is unobservable this assumption is untestable. Of course, this does not preclude the possibility of deriving a separate psychological time scale, as Micko (1969) proposes.

Now that we have dealt with the problem of what sort of scale RT lies on, we will turn briefly to consider two more advanced applications of measurement theory. The first involves the use of additive conjoint measurement to test whether two experimental factors additively affect RT. The second involves the technique of dimensional analysis. Although it is not quite clear what the ultimate impact of dimensional analysis will be in psychology, it is intriguing and seems well worth bringing to the attention of information processing theorists.

Additivity of RT and measurement theory

As noted, the question arises naturally as to the relation of RT additivity to additive conjoint measurement (Luce & Tukey 1964; Krantz et al. 1971). Although the possibility of such relationships has been anticipated, there appears to have been no work on this problem. As was seen earlier in this chapter, the main idea behind such treatments as Sternberg's (1969a, b) is that when random variables (e.g., processing times) are additive and separate functions of experimental factors, the expectations are also additive in those factors. There has as yet been little attempt to develop axiomatic measurement theories within a probabilistic framework; rather, the great bulk of the work has been based on deterministic variables. A recent exception is work by Falmagne (1976, 1979). For instance, in his 1976 paper, a theory of random conjoint measurement is applied to a loudness summation experiment and the measurement of median matching intensities within that context.

Now, medians do not necessarily add when the random variables themselves add, as do the means, so expectations seem preferable in investigations of additivity of processing times. In any event, the factor additivity question will be discussed here from the viewpoint of conjoint measurement without regard to the question of statistical error.

Suppose we have two sets, A_1 and A_2, which typically may refer to physical dimensions or factors like X_a and X_b, and we are interested in varying values from the two sets and seeing what happens on some dependent variable, such as RT. The notation follows closely that of Krantz et al. (1971), and the reader is referred to that work for further discussion and proofs. The idea in such developments as these is to establish empirical relationships that must hold among the members of A_1 and A_2 with regard to their effect on the

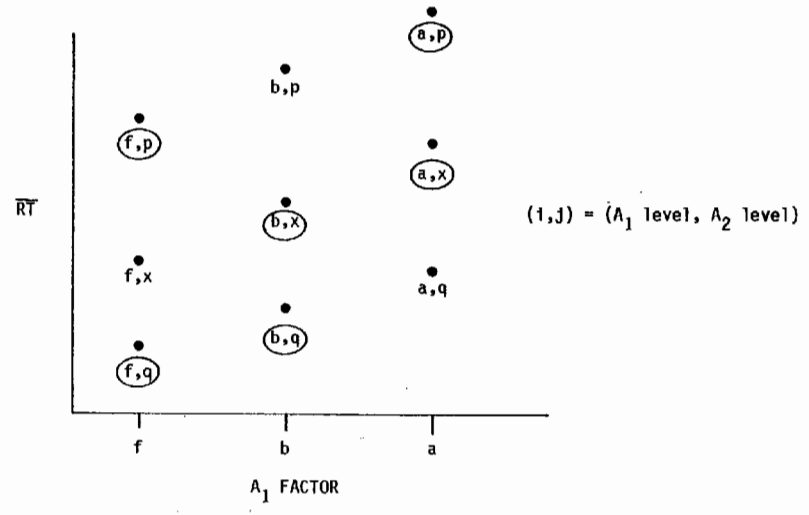


Fig. 12.8. A 3×3 factorial experiment. Circles indicate conditions sufficient to test double cancellation.

dependent variable in order that the dependent variable of interest will be an additive function of the members of A_1 and A_2.

Typically, the axioms of conjoint measurement theory are stated in terms of the empirical relation $\succsim$. Because the empirical relation is unobservable in the case of RT, we must state the axioms in terms of the numerical relation $\geqslant$. The psychological object of interest is $RT(a, p)$, a function of the ordered pair $(a, p) \in A_1 \times A_2$.

We proceed in our discussion by listing six axioms on the triple $\langle A_1, A_2, \geqslant \rangle$ that are jointly sufficient to guarantee additivity (see Krantz et al. 1971 for a proof of sufficiency). Thus if the six axioms hold we know $RT(a, p) \geqslant RT(b, q)$ if and only if $T_1(a) + T_2(p) \geqslant T_1(b) + T_2(q)$, where, for example, $T_1(a)$ is the duration of subsystem S_1 when factor X_a is at level a.

Because we are stating the axioms in terms of a numerical RT scale that we are assuming is of ratio level, it must satisfy the conditions of a weak order (Krantz et al. 1971) and so the first condition below could be dropped from our list. Nevertheless, it is recounted here for completeness.

Figure 12.8 illustrates a hypothetical experiment where A_1 = stimulus intensity, A_2 = response compatibility, and RT is a conjoint function of these. Note that perfect additivity occurs.

We now turn to the six axioms jointly sufficient for factor additivity.

Definition 12.1: The relation on $A_1 \times A_2$ is a *weak order* relation if and only if for any $a, b, c \in A_1$ and $p, q, r \in A_2$, it is

(i) *connected:* either $RT(a,p) \geqslant RT(b,q)$ or $RT(b,q) \geqslant RT(a,p)$

and

(ii) *transitive:* if $RT(a,p) \geqslant RT(b,q)$ and $RT(b,q) \geqslant RT(c,r)$ then $RT(a,p) \geqslant RT(c,r)$. □

As stated above, this condition follows trivially from our assumption that processing time is on a ratio scale.

The second definition establishes that each component (factor, e.g., X_a or X_b) can be varied without affecting the other.

Definition 12.2: The relation $\geqslant$ on $A_1 \times A_2$ is *independent* if and only if for any $a, b \in A_1$, $RT(a,p) \geqslant RT(b,p)$ for some $p \in A_2$ implies that $RT(a,q) \geqslant RT(b,q)$ for every $q \in A_2$; and for any $p, q \in A_2$, $RT(a,p) \geqslant RT(a,q)$ for some $a \in A_1$ implies that $RT(b,p) \geqslant RT(b,q)$ for every $b \in A_1$. □

Thus, we see that the ordering on RT forced by what corresponds to an inequality on one dimension has to continue to apply no matter what constant value is placed on the other dimension.

When independence holds, no crossovers can occur in the RT data [e.g., $RT(a_1,p_2) > RT(a_2,p_1) > RT(a_1,p_1) > RT(a_2,p_2)$ violates independence], and conversely, when crossovers unambiguously are in evidence, then independence is false. However, an absence of crossovers does not justify the conclusion of additivity, of course; other assumptions, in addition to independence and the weak order notion, must be satisfied. The next assumption, the double cancellation axiom, and its immediate consequent, the Thomsen condition, are critical aspects of conjoint measurement. Put verbally, the double cancellation condition says that if the effect of the pair (a,r) is greater than the effect of (c,q), but the pair (c,p) has a greater effect than (b,r), then certainly the pairing (a,p) should have a greater effect than the pair (b,q). This makes intuitive sense because we are selecting the "stronger" values a and p to act against the "weaker" values b and q. Interestingly, this condition does not follow from independence or the other axioms.

Definition 12.3: The relation $\geqslant$ on $A_1 \times A_2$ satisfies *double cancellation* if and only if, for every $a, b, c \in A_1$, and $p, q, r \in A_2$, when $RT(a,r) \geqslant RT(c,q)$ and $RT(c,p) \geqslant RT(b,r)$, then $RT(a,p) \geqslant RT(b,q)$. The weaker condition in which $\geqslant$ is replaced by $=$ is the *Thomsen condition.* □

An example might help clarify this very important assumption. Suppose A_1 is modes of transportation and A_2 is authors; and suppose further that mean reaction time $\overline{RT}$ is monotonically related to preference on pairs of transportation and author choices. That is, $\overline{RT}(a,p) \leqslant \overline{RT}(b,q)$ if and only if (1) the means of transportation conferred by a together with a book by author p is preferred at least as much as transportation b with author's book q, and (2) the mean RT to (a,p) is faster than to (b,q). Now let $\gtrsim$ refer to a weak preference ordering. Here, then, it is reasonable to think that on A_1 above, airplane $\gtrsim$ sportscar $\gtrsim$ motorcycle (at least for some individuals) and, on A_2, Dostoyevsky $\gtrsim$ Samuel Pepys $\gtrsim$ Harold Robbins. Let a = airplane, b = sportscar, and c = motorcycle, and p = Dostoyevsky, r = Samuel Pepys, and q = Harold Robbins. Then if, as seems plausible, airplane and Samuel Pepys $\gtrsim$ motorcycle and Harold Robbins, and motorcycle and Dostoyevsky $\gtrsim$ sportscar and Samuel Pepys, then it is only "logical" that airplane and Dostoyevsky $\gtrsim$ sportscar and Harold Robbins. Therefore, in terms of $\overline{RT}$, the conditions are

$$\overline{RT}(\text{airplane, Samuel Pepys}) \leqslant \overline{RT}(\text{motorcycle, Harold Robbins})$$
$$\overline{RT}(\text{motorcycle, Dostoyevsky}) \leqslant \overline{RT}(\text{sportscar, Samuel Pepys})$$

and, by double cancellation,

$$\overline{RT}(\text{airplane, Dostoyevsky}) \leqslant \overline{RT}(\text{sportscar, Harold Robbins})$$

Notice that Dostoyevsky was sufficiently strong even when paired with motorcycle to outweigh the combination of moderate strength, sportscar and Samuel Pepys. It is therefore not surprising that when combined with airplane it was faster (and preferred to) sportscar with Harold Robbins. Note also that airplane and Samuel Pepys *must* be stronger than motorcycle and Harold Robbins, once we know the separate dimensional orderings by independence (Definition 12.2).

The next definition captures the fourth assumption that is necessary for additivity (as are the foregoing three). It is what is often referred to as a *technical axiom,* because the ultimate mathematical solution depends on it, but it is not easy to see how it would be tested empirically. It is called the *Archimedean axiom* and employs the notion of a standard sequence. A standard sequence on A_1 is formed by selecting p and q, with $RT(a,p) \neq RT(a,q)$ (i.e., p is not of identical strength to q in terms of RT) on A_2 and then finding a set $\{a_i \mid a_i \in A_1$, for $i \in$ set of consecutive integers$\}$ such that $RT(a_i,p) = RT(a_{i+1},q)$ for $i = 1, 2, \ldots$. An analogous definition holds for the other dimension.

For a better intuitive understanding of a standard sequence suppose additivity holds. Then

$$RT(a_i,p) = RT(a_{i+1},q)$$

implies

$$T_1(a_i) + T_2(p) = T_1(a_{i+1}) + T_2(q)$$

or equivalently

$$T_1(a_{i+1}) - T_1(a_i) = T_2(p) - T_2(q)$$

Similarly,

$$T_1(a_i) - T_1(a_{i-1}) = T_2(p) - T_2(q)$$

and thus

$$T_1(a_{i+1}) - T_1(a_i) = T_1(a_i) - T_1(a_{i-1}) = T_2(p) - T_2(q)$$

Thus, a standard sequence sets up intervals on A_1 that have equal effects on RT and whose magnitude is prescribed by the effects of the constant p-to-q difference.

Definition 12.4: A standard sequence is *Archimedean* if and only if whenever it is strictly bounded, it is finite. □

This axiom is pesky, rather like the axiom of choice, which assumes that given any collection of sets, each with at least one member, there exists a selector function that can pick a member from each of the sets, even though the individual sets and even the class of sets may be of infinite number (not to be confused with Luce's choice axiom, 1959a). Fastidious (perhaps *prudent* would be a less pejorative term) mathematicians would prefer to do without the axiom of choice, but cannot without risking the loss of many important results. We shall not comment further on the Archimedean axiom here.

Solvability, in RT applications, means the ability to find, for example, the missing component in $\mathrm{RT}(a,p) = \mathrm{RT}(?,q)$, when the other three are given. Unrestricted solvability implies that this is true for any triple a, p, and q, whereas restricted solvability only stipulates that a solution need occur if p and q are not both so near an upper or lower bound that no b can be found to go with an equally extreme a. We shall only give the second definition here, because it is all we require.

Definition 12.5: A relation on $A_1 \times A_2$ satisfies *restricted solvability* provided that (i) whenever there exist $a, \bar{b}, \underline{b} \in A_1$ and $p, q \in A_2$, for which $\overline{\mathrm{RT}}(\bar{b}, q) \geqslant \overline{\mathrm{RT}}(a,p) \geqslant \overline{\mathrm{RT}}(\underline{b}, q)$, then there exists a b contained in A_1 such that $\overline{\mathrm{RT}}(a,p) = \overline{\mathrm{RT}}(b,q)$; (ii) a similar condition holds on the second component. □

The last definition simply states that each of the dimensions A_1 and A_2 actually have some effect on the dependent variable.

Definition 12.6: Suppose that $\geqslant$ is a relation $A_1 \times A_2$. Component A_1 is *essential* if and only if there exist $a, b \in A_1$ and $p \in A_2$ such that $\overline{\mathrm{RT}}(a,p) \neq \overline{\mathrm{RT}}(b,p)$. A similar definition holds for A_2. □

The next proposition states that an additive representation exists if the above six definitions are assumed. These conditions on the triple $\langle A_1, A_2, \geqslant \rangle$ result in that triple being what is called an *additive conjoint structure*.

Proposition 12.7: Suppose $\langle A_1, A_2, \geqslant \rangle$ is an additive conjoint structure in RT. Then there exist functions T_i from A_i, $i = 1, 2$, into the real numbers such that, for all $a, b \in A_1$ and $p, q \in A_2$, $\mathrm{RT}(a,p) \geqslant \mathrm{RT}(b,q)$ if and only if $T_1(a) + T_2(p) \geqslant T_1(b) + T_2(q)$. If T_i' $(i = 1, 2)$ are two other functions with the same property, then there exist constants $\alpha > 0$, and β_1 and β_2, such that $T_1' = \alpha T_1 + \beta_1$ and $T_2' = \alpha T_2 + \beta_2$. Thus, we are ensured a numerical additive representation of RT and that any two differing such representations are related by a linear transformation.

Proof: (See Krantz et al. 1971.) □

How does all this apply to RT? Well, of course, it cannot be applied to empirical RTs as stochastic entities without some intrusion of probabilistic concepts into (or instead of) the above axioms. On the most fundamental level it may be preferable to approach the problem more from the stochastic modeling point of view as we have elsewhere in the book. Perhaps as knowledge grows concerning probabilistic fundamental measurement, synthesis of measurement-oriented and stochastic information process-oriented approaches will appear.

However that may be, we will proceed with our discussion by assuming that we are interested in establishing an additive representation for $\overline{\mathrm{RT}}$. The question may be asked whether the potential additivity of $\overline{\mathrm{RT}}$ can be tested by observing whether those assumptions with some empirical bite hold in a set of data. Obviously, the most important are the necessary axioms, because if one of them is falsified, no additive representation can exist. As we noted earlier, the weak order condition is automatically satisfied by $\overline{\mathrm{RT}}$. The independence definition (Definition 12.1) might be falsified in some cases, and if so will rule out additivity of the means. But the most critical is probably the double cancellation axiom (Definition 12.3), which is easier to test in the usual RT experiment than would be the Thomsen condition, because the latter requires exact settings of A_1 and A_2 values in order to yield equal RT magnitudes (which is why we used the double cancellation condition rather than the Thomsen condition in our assumptions even though only the Thomsen condition is required with our other assumptions).

Double cancellation could be tested by selecting magnitudes a, b, and c in A_1 and p, q, and r in A_2 such that $\overline{\mathrm{RT}}(a,r) \geqslant \overline{\mathrm{RT}}(c,q)$ and $\overline{\mathrm{RT}}(c,p) \geqslant \overline{\mathrm{RT}}(b,r)$ and then by checking whether $\overline{\mathrm{RT}}(a,p) \geqslant \overline{\mathrm{RT}}(b,q)$. Notice that we have three values or levels on each of A_1 and A_2. It is not necessary to experimentally present all nine combinations, though, in order to check the double cancellation axiom, although the latter will allow for more checks. Figure 12.8 shows an example of a 3×3 factorial experiment with circles drawn around six pairs that suffice to test double cancellation. Naturally, in real data, the means will be estimates of the true means and subject to error. Because their standard deviations will also be estimable, common statistical tests can be performed.

On the other hand, why should we not simply perform the slightly extra experimentation necessary to obtain all the points in Fig. 12.8 and test additivity by analysis of variance or some similar technique? As we stated earlier, the axioms of additive conjoint measurement are usually stated in terms of

the empirical relation $\gtrsim$. This obviates the need to specify the exact nature of the relevant numerical scales, and herein lies the true power of the theory. Analysis of variance, on the other hand, provides a method for testing additivity on numerical data and therefore requires knowledge of the true underlying numerical scales. Thus if one finds additivity with analysis of variance (assuming a reasonable level of power and that the appropriate statistical assumptions are satisfied), the axioms of additive conjoint measurement should be satisfied.

On the other hand, if an interaction is found, the conjoint measurement axioms may or may not be satisfied. For example, if additivity holds but the wrong numerical scale is used, it is possible that an interaction will be found with analysis of variance even though the conjoint measurement axioms all hold. In this case, some transformation of the employed scales can be found that will yield an additive numerical representation.

In the case of RT, however, the conjoint measurement axioms are stated in terms of the same numerical scale that the analysis of variance would be performed on. If we believe that RT is the correct underlying scale and if we assume it lies on a ratio scale, then the only admissible transformations on the numerical RTs are similarity transformations (multiplication by a positive constant). But because similarity transformations map parallel curves into parallel curves, an interaction found with analysis of variance implies that no additive representation of RT exists. This is the same conclusion we would draw if one or more of the conjoint measurement axioms were violated when they are stated in terms of RT. Thus, in the case of RT, additive conjoint measurement and analysis of variance appear to provide equivalent tests of additivity. Of course, in empirical applications, because the two techniques make rather different statistical assumptions, one would expect to occasionally find instances in which they disagree.

The primary advantage of testing additivity with conjoint measurement rather than analysis of variance is that it does not require knowledge of the true underlying numerical scales. The primary disadvantage is that, because it is a deterministic theory, conjoint measurement provides no basis for statistically testing its predictions. Unfortunately, in the case of RT, this advantage is lost, whereas the disadvantage remains. On the other hand, analysis of variance does make a number of statistical assumptions. Nevertheless, until some technique is developed for observing the empirical processing time relation, it appears that, as a test of RT additivity, conjoint measurement has few advantages over more traditional techniques such as analysis of variance.

Dimensional analysis

A final measurement issue is related in a general way to the foregoing thoughts. Although the results of the following theory apparently cannot be presently used in psychology, it is of considerable scientific and philosophical interest. It also serves to point up the chasm that currently separates the physical from the nonphysical disciplines. If a functional relation holds between two variables x and y, and their scale types (e.g., ratio, ordinal) are specified, are there any restrictions imposed by the scale types on the form of the functional relationship? Luce (1959b) proposed that an admissible transformation on the argument of the function x (the independent variable) should lead only to an admissible transformation on the scale values of the resulting y (the dependent variable). Further, Luce also assumed that the functional form of the law relating x and y should not depend on the particular scales chosen to measure x and y. He then showed that these stipulations permitted the formulation of functional equations that determined the possible forms of the functional relationship holding between x and y. Rozeboom (1962) questioned the independence of the two conditions.

In Luce's (1959b) notation, suppose $y=u(x)$. If both x and y are ratio scales, then the only admissible transformation on either is multiplication by a positive constant. The major thrust of imposing the above constraint on u is the functional equation $u(kx)=K(k)u(x)$, where k is the positive constant multiplying x (thereby changing its unit) and $K(k)$ is the comparable and subsequent change of the unit of y and will typically be, as shown, a function of the change of unit (k) of x. Note that it would be incorrect to write, $u(kx)=K(k)u(x)+C(k)$, because that would amount to a nonadmissible transformation on the dependent variable y. However, in the event that y was an interval scale, with x still a ratio scale, then the latter would be the proper functional equation (see Luce 1959b: Table 1).

Solving the former functional equation (with both x and y as ratio scales and assuming continuity or monotonicity on an interval) yields $u(x)=\alpha x^{\beta}$ as the only possible form that u can take. For instance, changing x to kx (e.g., meters to centimeters) then gives $y=u(kx)=\alpha k^{\beta}x^{\beta}=\alpha' u(x)$, where $\alpha'=K(k)=k^{\beta}$, the change of unit in y. Hence, we see that the only possible function relating x and y is a power function.

One problem in applying the method is that in certain types of laws, the scales (dimensions) cancel out, in which case no constraints are imposed on the functional law relating x and y (Luce 1959b, 1962; Rozeboom 1962). We illustrate this below, but first we would do well to place the discussion in a somewhat broader perspective.

We have been proceeding within the context of the treatise by Luce (1959b). Krantz et al. (1971: Chapter 10) provide a much deeper penetration of the interrelationship of laws in physics via the measurements underlying those laws. A good deal of follow-up work has appeared more recently (Cohen & Narens 1980; Luce 1978; Narens 1981; Narens & Luce 1976). In fact, the later works proceed along a rather different path, one closely related to the classical theory of dimensional analysis in physics. One outcome of the development by Luce and his colleagues is a more rigorous underpinning of dimensional analysis within the context of measurement theory. (A readable introduction to the practical uses of dimensional analysis can be found in Huntley 1967; Sedov 1959 offers many interesting examples.)

In order to illustrate when dimensional analysis is helpful and when not, consider the extremely simple example of a spring. Suppose it has been observed (by a being relatively unversed in earthly mechanics) that a given spring responds with more force the more distended it is, but that for the same amount of force, two different springs exhibit different lengths of stretch. Suppose the being happened to conjecture that the force exerted by a spring is associated with some distension that is multiplied by a constant k that characterizes the particular spring. An elementary problem is then to describe the units of the "spring constant" that multiplies the length of distension. One might, according to the strategy of dimensional analysis, write down the dimensions on the two sides of a dimensional equation relating length and force as follows:

$$F=D(k)L \quad \text{or} \quad MLT^{-2}=D(k)L$$

where L=length dimension in ft, F=force dimension, M=mass dimension in lb-sec^2/ft, T=time dimension in sec, and $D(k)$=dimension of spring constant. It can be seen that we are taking length, mass, and time as basic, and other dimensions like force as secondary. It is important to note that the above equation concerns only the dimensions, *not* particular quantities along those dimensions, relative to some set of units. The latter will be designated with small letters, for example, ℓ=length quantity, and so on. Clearly, the overall dimensions of the two sides of the equation have to match up, so the dimension of the spring constant $D(k)$ must be $D(k)=MT^{-2}$, that is, mass divided by time-squared. This is a relatively trivial use of dimensional analysis. We will shortly give a dynamic view of the same types of system in which the relevant dimensions cancel out and thus do not directly determine the functional relationship. First, however, as a slightly more interesting example, consider a simple pendulum of length ℓ and bob mass m and suppose it is hypothesized that the period of the swing is some function of the mass, the length, and the weight of the bob (the latter is in force units). Assuming that such physical laws are expressed by multiplication and division supplemented by a possible dimensionless constant (Krantz et al. 1971 and later work spend much time providing justification for this), it follows that in terms of dimensions, the period dimension (T) can be written

$$T=M^aF^bL^c=M^a(LMT^{-2})^bL^c$$

where the a, b, c are exponents to be determined. The constraint that the right-hand dimensions resolve down to T means that the linear equations $a+b=0$, $b+c=0$, and $-2b=1$ must hold, and these are satisfied if $b=-\frac{1}{2}$, $c=\frac{1}{2}=a$. This implies that the law written in terms of physical quantities (numbers reflecting the magnitudes along the various dimensions) can be expressed as

$$t=km^{1/2}(f)^{-1/2}\ell^{1/2}=k\left(\frac{m\ell}{f}\right)^{1/2}$$

where m=mass, f=force, ℓ=length, and k=dimensional constant. This expression may be simplified further by noting that m/f corresponds to the dimensions $M/(MLT^{-2})=1/(LT^{-2})$, where LT^{-2} comprises the dimensions of gravity g so that $t=k(\ell/g)^{1/2}$, the usual expression of the law in terms of length and the gravitational constant of the particular planetary body on which the pendulum swings.

Let us return to the spring example in order to show how dynamical laws may appear in a dimensionless fashion (a similar example could be constructed with the pendulum).

To keep matters simple, assume that once the spring has been displaced to length e, it is released and under free motion. We then write the force equation in a way that exhibits the motion (displacement position over time). Again, let m=mass, ℓ=displacement in ft, t=time in sec, and k=the spring constant in mass/sec^2. Then the equation expressing the spring motion over time is

$$m\frac{d^2}{dt^2}\ell=-k\ell$$

This is like the usual statement "the force acting against the spring is inversely proportional to the displacement of the spring." Solving this for $\ell(t)$ in terms of t finds

$$\ell(t)=\ell_0\sin\left(\frac{\sqrt{k}}{\sqrt{m}}t+\frac{\pi}{2}\right), \quad \ell_0=\text{initial displacement}$$

a form of simple harmonic motion (note that the deceleration factor has been omitted; the conclusions would be the same with regard to dimensional analysis had it been included). Now the dimensions of $\sqrt{k}t/\sqrt{m}$ are

$$\frac{(MT^{-2})^{1/2}}{M^{1/2}}T=1$$

so that the dimension of the independent variable t cancels out, as does the mass dimension. This result does not contradict dimensional analysis; in fact, it is in accord with it because only the dimension of length (ℓ_0 and ℓ) remains on the right-hand and left-hand sides. The point, as earlier expressed by Rozeboom (1962) and Luce (1962), is that the dimensions are unrelated here to the dynamic activity of the system across time, due to the initial cancellation.

It may be asked what relationship is borne between dimensional analysis as practiced above and the aforementioned early functional equation work by Luce (1959b). Dimensional analysis, as is apparent above, *assumes* that physical scales (which are ratio and log interval) appear in physical laws as power functions (with rational exponents) of one another. The approach taken by Luce, on the other hand, is one means, no longer considered the best one, of justifying the power function relationship (see also Krantz et al. 1971: Sections 10.7, 10.9, 10.12).

What is the impact of this theory for psychology? Unfortunately, so far, not very much. The major results do not generalize well to interval and

weaker scales, and the behavioral and social sciences are not presently overburdened with ratio scale measurement.

Krantz et al. (1971) suggest that the nonphysical sciences must

1. discover their own ratio scales and append them to the existing structure of physical quantities, or
2. introduce into the structure new nonbasic quantities that are basic to the nonphysical sciences, or
3. arrive at lawful formulations having a different character from the dimensionally invariant equations of physics.

The first alternative seems presently hopeless at least for large regions of psychological investigation. The second shows some promise in psychophysics where scaling of psychological attributes in terms of physical quantities has a relatively long and fruitful history; Krantz et al. show how such approaches as Stevens (1959, 1966) took might be interleaved with classical dimensional approaches to arrive at the physical dimensions corresponding to such psychological attributes as subjective intensity.

The third possibility refers primarily to (a) developing laws that violate classical dimensional laws or (b) avoiding such difficulties as those encountered with interval scales by performing manipulations such as taking ratios of differences (which essentially converts interval to ratio scales).

Two other potential approaches come to mind. First, it seems possible at least in principle that when the nervous system and its metabolism and possibly the neuroendocrine systems are better understood, for certain delimited psychological behaviors a particular neural subsystem might be approximated as a closed system over the limited period of the behavior. Then it might be feasible to employ a reductionistic argument in relating the ordinary dimensions of physics and chemistry to link the independent and dependent variables to one another, as accomplished in traditional applications of dimensional analysis.

The other approach basically reflects the status quo, namely, virtually all laws of the social sciences and many of the biological sciences do not violate dimensional theory; they simply do not interface with that theory. It might be argued that the behavioral sciences have come along just fine without attempting to perfectly imitate physics and that, in fact, it may be a waste of time and energy to do so.

The connecting fibers between this discussion and stochastic processing theory are admittedly slight at this time. However, if and when dimensional analysis can be amalgamated into the nonphysical sciences, RT should be one prime candidate as a dependent variable, particularly if, as our former arguments suggest, it can be taken as a ratio scale.

Thus, suppose such an amalgamation takes place with regard to physical variables such as stimulus energy and certain dependent measures such as RT and that all of these lie on ratio scales. Then the relation between, say, stimulus energy and RT is a power function with a rational exponent, and the exponent itself should be determinable via dimensional analysis. One would

imagine, too, that many dynamic accounts, for instance, the spatial position of the hand in a tracking task, will obey the dimensionless account, where the dimensions cancel out and are of little aid in determining the overall systemic law (as was the case in the dynamic model of the single-spring system above).

An introduction to systems and automata theory in relation to additivity of reaction times

Many behavioral and neural scientists have employed analogies borrowed from what we may loosely call *systems theory,* and this propensity is increasing every year. Perhaps the least constrictive analogy (with the attendant advantages as well as pitfalls) is so-called *general systems theory* (there are a number of books on this, e.g., Klir 1969), which imposes few restraints on the workings of the system from the outset. Somewhat more restricted is *automata theory* (Booth 1967; Starke 1972), with finite automata as a special case. *Linear systems theory* (Padulo & Arbib 1974; McGillem & Cooper 1974; Norman 1981) is quite restrictive but has been of enormous importance in the other sciences and in psychoacoustics and to a lesser extent in visual psychophysics as well as in modeling of simple motor-tracking behavior. In fact, when the linearity assumption is dropped, matters get quite difficult, and it cannot be said that nonlinear systems theory has yet had the impact that might have been hoped for, due to its mathematically refractory nature.

Virtually all reasonably defined systems can be put into the framework of a set U of inputs, a set of stages, Q, associated with the internal condition or states of the system, and a set Y of outputs. The states and outputs change as a function of processes and present states and inputs, and, in the most general cases, with time as well. When the changes of states are probabilistic rather than deterministic, we have the notion of stochastic systems. The large part of the work of this volume is, of course, an investigation of certain classes of systems. As before, when the distinction is important, we employ *system* to refer to the actual "real-life" entity doing the processing and *model* as the mathematical description of a system.

In order to provide a modicum of background to systems theory, a relatively informal introduction to deterministic linear systems is given, and then rather evident sufficient constraints that lead to nonoverlapping serial systems or to the less restrictive mean RT additivity are proposed. Then a fairly general description of a class of stochastic automata systems is offered of which discrete linear systems are a special case. Finally, the conditions under which such systems produce strictly serial behavior and mean RT additivity are stated.

Linear systems

Unlike the typical RT analysis as employed by psychologists, linear systems theory is based first and foremost on the input and output and state

of the system rather than the time taken by the system. Certainly temporal phenomena immediately ensue from the input–output characteristics of a system, but they are not the only matter of concern. Psychology typically has little evidence concerning how neural or cognitive processing is associated with the input and output of psychological functions and so has tended to focus, when attempting to analyze and identify separate subsystems, on time properties of the human processing system.

It will be convenient to develop the simple case where the set of states Q equals the set of outputs Y and thus an output uniquely defines the state of the system at any point in time. Here we shall use the idea of a linear, causal, time-invariant system. *Linear* simply means that if the response of the system to input $u_1(t) \in U$ is $y(t) = L[u_1(t)]$ and to input $u_2(t) \in U$ is $y(t) = L[u_2(t)]$, then the response to the combined input $u_1(t) + u_2(t)$ is just

$$y(t) = L[u_1(t) + u_2(t)] = L[u_1(t)] + L[u_2(t)]$$

that is, the output is the sum of the two outputs associated respectively with the separate inputs. This property of superposition must hold, of course, for all time t considered.

Causal systems are ones where the output cannot be affected by future inputs, and time invariance implies that shifting the input by any fixed time T only shifts the output accordingly. Intuitively, the internal processing characteristics of the system are not varying over time, for instance, through fatigue, warmup effects, or the like. Systems such as these may be fruitfully studied by emphasizing the time domain and the state of the system (including the input and output). This approach centers on mappings from the input to the system state and thence to the output (this latter mapping is the identity in our $Q = Y$ example), and then develops differential or integral equations to relate the input and output. Another method involves taking a Fourier or Laplace transform of the time domain functions and concentrates on the relative amplification of the various frequencies and phase shifts in the input signal by the system under study.

More formally, let the sets U and $Q = Y$ both be the real-line $\mathbf{R}$.

An integral equation may be written that follows the above precepts. Let $h(t) \in \mathbf{R}^+$ be a linear weighting function that gives the amplification or weight that is accorded the input that actually occurred t time units in the past. Thus, $h(0)$ is the amplification given the present input and if the input started back at $t = 0$; then $h(t)$ reveals the importance at the present moment of the input that happened t time units ago. For the system to be *causal*, $h(-t)$ must not be allowed to be other than 0, for otherwise the system would be weighting "inputs" from the future. Any instant t', then, from the past to the present, contributes an amount $u(t')h(t-t')$ to the present state of the system at time t $(t' \leqslant t)$. As noted above, $y(t)$ is both the state of the system and also the output function. This is, of course, a special case of the general situation where the state of the system is then mapped into the output (put another way, the output mapping is the trivial multiplication by 1). We

assume that the input u begins specifically at $t = 0$, but this is of little consequence since the system is time-invariant. *Time invariance* means that the weighting function $h(\cdot)$ does not depend on the exact present time, only on the time lapsed between an input and the present. In the general time-variant case, we would have to write $h(t', t)$, with t representing the present and t' the past instant of input.

We can now obtain the "total" output at the present instant t by integrating all of the contributions from time 0 up to time t. It will be assumed in Eqs. 12.11–12.18, for illustration, that t, the time of measurement of the output, and the various $(t-t')$ terms do not exceed the duration of the input or the lengths of time that h_a and h_b are nonzero. This assumption will be relaxed later. Then

$$y(t) = \int_0^t h(t-t')u(t')\,dt' \tag{12.11}$$

Here, we identify y both as the state of the system and as the output function.

It may be noted that by convention we let $h(0)$ be the weighting at the present where we have $h(0)u(t)$ to represent the system output at that moment and to run the independent variable or argument of h upward in positive value to stand for increasing times *in the past*. Thus, if we graph the function h we must keep in mind that we multiply the input u by $h(t-t')$ rather than by $h(t)$; that is, we can think of the weighting process as running h backwards relative to the time progression of $u(t)$. In any case, the expression Eq. 12.11 is actually symmetric in the sense that it can be written equivalently as

$$y(t) = \int_0^t h(t')u(t-t')\,dt'$$

This type of mathematical operation is known as *convolution* (see Chapter 3). It was employed in earlier chapters where it was associated with the probability density function of the sum of two independent random variables. Here, however, it represents the output of a linear system. It, along with correlation, is also important in memory storage and retrieval and recognition theories (e.g., Borsellino & Poggio 1973; Gabor 1968, 1969; Hinton & Anderson 1981; Murdock 1979; Pribram 1966).

Figure 12.9a–d shows a graphical illustration of the convolution operation in which the weighting function is exponentially damped and the input is a square wave. Note that the weighting function assigns weight to all inputs that have ever been received by the system, but that the farther back in time the input occurred the less weight it is assigned and hence the less it contributes to the output of the system. It can be seen that the output of the system $y(t)$ increases exponentially as long as the stimulus is present (i.e., as long as the input is nonzero), and when the input is terminated the response of the system begins to immediately die out.

The discrete analog to Eq. 12.11 is given by

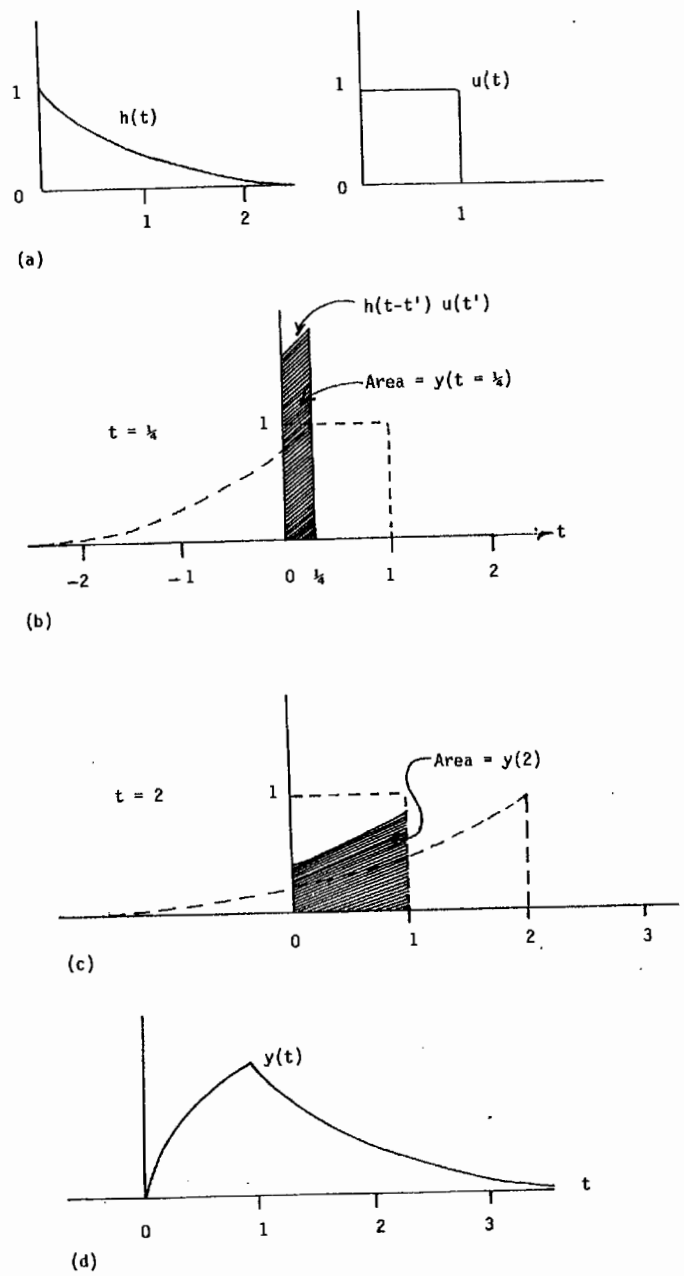


Fig. 12.9. (a) The manner in which a linear time-invariant causal system weights the input; the input here is $u(t)$=square wave, and the system-weighting function (the so-called impulse response function) is $h(t)$=exponential. Input and weighting function are shown separately. (b) The shaded trapezoid gives in height [height$=h(t-t')\cdot u(t')$] the contributions to the output y for t between $t=0$ and $t=\frac{1}{4}$. The shaded area equals the actual magnitude of the output y at $t=\frac{1}{4}$, i.e.,

$$y(\tfrac{1}{4})=\int_0^{1/4}\exp[-(\tfrac{1}{4}-t')]\,dt'=1-e^{-1/4}\simeq .22$$

In (c) the input ceases at $t=1$ so that at $t=0$, $h(t-t')$ is multiplying 0 for $t'=1$. The output is now

$$y(2)=\int_0^2 e^{-(2-t')}\,dt'$$

$$=\int_0^1 e^{-(2-t')}\,dt'=e^{-1}-e^{-2}\simeq .50$$

In (d) the continuous successive output of $y(t)$; note the decay beginning with the offset of the input.

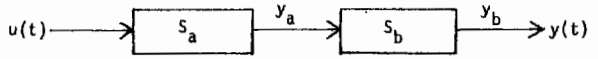


Fig. 12.10. A series connection of subsystems.

$$y(k\Delta t)=\sum_{i=0}^{k} h(k\Delta t-i\Delta t)u(i\Delta t) \tag{12.12}$$

In Eq. 12.12, time can, of course, be thought of as separated into time quanta Δt, each Δt being worth a fixed number of time units. Clearly, $k\Delta t=t$ and $i\Delta t=t'$ in the continuous formulation.

Now consider what happens when two subsystems are connected in series, as in Figure 12.10. We shall use the term *series* as distinguished from *serial* by permitting subsystems connected in series to overlap in their processing times. Let us call the output of subsystem S_b, y_b. Then the output y_a becomes the input to S_b, of course. Note that we assume the transmission time to S_b, given that we have y_a at hand, is instantaneous. This can easily be weakened by imposing a delay mechanism between S_a and S_b, if necessary. For neural subsystems juxtaposed in the sense of being connected through a set of immediate synapses, the outputs of S_a terminal endbulbs (the output channels) would feed into the S_b dendrites (the S_b input channels) with a delay that is small relative to the usual time course of events.

We thus have constructed the system S composed of

$$y_a(t')=\int_0^{t'} h_a(t'-t'')u(t'')\,dt''$$

$$y_b(t)=\int_0^{t} h_b(t-t')y_a(t')\,dt' \tag{12.13}$$

that is,

$$y_b(t)=\int_0^t h_b(t-t')\left[\int_0^{t'} h_a(t'-t'')u(t'')\,dt''\right]dt'$$

which is the convolution of the subsidiary output y_a with the S_b weighting function. Thus, using our Chapter 3 notation,

$$y_b(t)=h_b(t)*[h_a(t)*u(t)] \tag{12.14}$$

However, because convolution is an associative operation, this expression may be rewritten in a more revealing fashion as

$$y_b(t)=[h_b(t)*h_a(t)]*u(t) \tag{12.15}$$

This result is derived in the discrete case in the appendix to this chapter.

The output y_b is therefore equal to the result of convolving the input u with the convolution of h_a and h_b. This suggests that an equivalent system would result if, instead of passing u first into S_a and the result of that into S_b, we simply began with a system possessing the weighting function h_a*h_b. In fact, this is true; such systems are completely equivalent.

We saw in Chapter 3 that the characteristic function (or moment-generating function when it exists) of the sum of two independent random variables is simply the product of the two separate characteristic functions. Because the probability density of such a sum is the *convolution* of the separate probability densities, it is not hard to reckon that the Fourier transform of the system composed of h_a*h_b will be equal to the product of the respective Fourier transforms. The Fourier (or Laplace) transform of any weighting function h, called the system or transfer function, gives a complete description of the system, and is denoted $H(f)$, where f refers to frequency. For example, in the case of two subsystems arranged in series, the transfer function of the composite system is $H(f)=H_a(f)H_b(f)$. We should note in passing that any $H(f)$ depicts the system in terms of the way that it handles the various frequencies embedded in the input and that furthermore $|H(f)|^2$ (the absolute value of the transfer function squared) represents the power allocated by the system specifically to amplify frequency f (see also Townsend & Ashby 1978 with regard to power interpreted as capacity).

We will assume that a response is given by such a system as soon as processing is completed – that is, when the output of the last subsystem becomes zero. In this case sufficient conditions for Donderian processing to occur are that (a) the processing operations or subsystems are arranged serially in the sense defined in earlier chapters, and (b) the experimenter can eliminate or add one of the subsystems by reducing or supplementing task complexity. This permits the method of subtraction to be employed. Sufficient conditions for mean RT ($\overline{\text{RT}}$) additivity, on the other hand, are weaker because it need only be assumed that the subsystems are hooked up serially and that different experimental factors separately affect distinct subsystems.

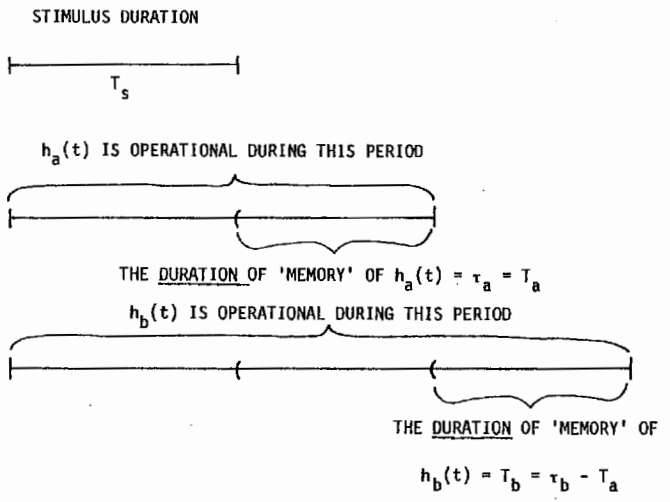


Fig. 12.11. Impulse response functions with cutoffs in time and the effects on serial subsystems S_a and S_b.

Notice that when h_a and h_b have sharp cutoffs in time [i.e., $h_a(t), h_b(t)=0$ for t greater than some values], the entire processing time may be expressed as in Fig. 12.11, yielding

$$T=T_s+T_a+T_b \tag{12.16}$$

Thus, obviously in this case if the cutoffs T_a and T_b are functions only of factors x_a and x_b, respectively, then RT will be an additive function of these factors. The next proposition makes this notation more precise.

Proposition 12.8: Assume as in Fig. 12.11 that $h_i(t)>0$ until time T_i where $i=a$ or b and that $h_i(t)=0$ after time T_i. Suppose T_a is a continuous decreasing function of x_a and T_b of x_b in that $T_a=T_a(x_a)$, $T_b=T_b(x_b)$ and that T_a and T_b are not functions of the alternate factor. Then

$$\text{RT}=T_a(x_a)+T_b(x_b)+T_s$$

Proof: This is reasonably obvious from Fig. 12.11. □

To get an idea of how a linear system acts in accordance with Proposition 12.8 in a simple case, suppose $u(t)=A\delta(0)$, that is, the input is an impulse (an instantaneous spike) of height $A>0$ occurring at $t=0$. Then the output will be nonzero up until $t^*=T_a(x_a)+T_b(x_b)$ and from that moment on will be 0.

In fact there will be a discontinuity at $t=t^*$ with the output at the instant $t^*=T(x_a)+T(x_b)$ being

$$\lim_{t\to t^{*-}} y_b(t) = \lim_{t\to t^{*-}} \int_0^t h_b(t-t') \int_0^{t'} h_a(t'-t'')u(t'')\,dt''\,dt'$$

$$= \lim_{t\to t^{*-}} \int_0^t h_b(t-t')h_a(t')\,dt' = h_a(T_a)h_b(T_b)$$

and

$$\lim_{t\to t^{*+}} y_b(t) = \lim_{t\to t^{*+}} \int_0^t h_b(t-t')h_a(t')\,dt' = 0$$

This result follows because to be nonzero, the argument in the integral must be such that $t' \leqslant T_a(x_a)$ and also $t-t' \leqslant T_b(x_b)$ so that $t-T_b(x_b) \leqslant t' \leqslant T_a(x_a)$. By symmetry of the convolution integral as well, $t-T_a(x_a) \leqslant t' \leqslant T_b(x_b)$. Thus, when $t=T_a(x_a)+T_b(x_b)+r$, $r>0$, then $r+T_a(x_a) \leqslant t' \leqslant T_a(x_a)$ and $r+T_b(x_b) \leqslant t' \leqslant T_b(x_b)$, which is impossible, but at the instant $t^*=T_a(x_a)+T_b(x_b)$ there is an output representing the respective magnitudes of $h_a[T_a(x_a)]$ and $h_b[T_b(x_b)]$.

Proposition 12.9 states conditions sufficient to produce a Donderian system – that is, one in which subsystem S_b becomes active only when processing by S_a is complete.

Proposition 12.9: A linear systems model, composed of two subsystems S_a and S_b in series, is a Donderian system if $h_a(t)>0$ only for $T_s \leqslant t \leqslant T_a$ and if $h_b(t)>0$ only for $T_a+T_s \leqslant t \leqslant T_b$, where T_s is the stimulus duration. □

The proof is straightforward and will not be given here. Thus, the conditions for a system to be Donderian are much stronger than they are for RT additivity. Not only must the impulse response functions have both an upper and a lower cutoff, but the lower cutoff of h_b must be precisely T_a+T_s.

We should mention some recent related work by McClelland (1979; see also Ashby 1982b), where a particular type of linear system is employed in which the weighting function $h_i(t)$ at each stage is a simple exponentially damped time function. A decision is made when a unit (subsystem) associated with a possible response at the highest level attains a criterion magnitude of activation (the criterion is allowed to vary to introduce some probabilism into an otherwise deterministic system). The characteristic of this system is used to investigate several experimental paradigms popular within cognitive psychology. Of most interest to the present discussion is McClelland's finding that when experimental manipulations affect certain parameters in separate serial subsystems, the appearance of temporal additivity is seen, whereas perturbation of other parameters evokes an apparent interaction. Conversely, experimental contrivances affecting the same subsystem can evince additivity or interaction depending on which parameters they alter.

On the other hand, some caution must be maintained when interpreting these results because Ashby (1982b) showed that this model does not have a well-defined RT distribution function because it always predicts a nonzero

probability that a response never occurs. By conditioning on the event that a response does occur, Ashby did derive RT density and distribution functions for McClelland's model, thus allowing most RT statistics to be computed directly and therefore eliminating the need for computer simulations. Using these results, an investigation of the model showed that it has severe problems predicting standard RT variability results because it predicts almost no increase in RT variability as the mean increases.

Finite-state probabilistic automata

Interestingly, when we move to a more complex type of system, the Donderian quality does not seem so bizarre, although mean RT additivity now appears rather strange; to be sure, there is a large subjective element in such an evaluation.

To keep matters simple, only finite-state automata operating in discrete times are considered, although they are permitted to be probabilistic. The discrete-time and finite-state assumptions obviate many analytic problems, although the logic of the arguments does not depend on the assumptions.

The notation required is listed below:

$U=\{u_j\}$ = finite set of n_u inputs to S_a.
$Q_i=\{q_j^{(i)}\}$ $(i=a,b)$ = set of n_{qi} states of system S_a or S_b.
$Y_i=\{y_k^{(i)}\}$ $(i=a,b)$ = finite set of n_{yi} outputs; the outputs of S_a serve as inputs to S_b, associated respectively with S_a or S_b.
Δt = basic time unit during which an input, output, and change of state occur; we can assume for simplicity that $\Delta t=1$.
$P_i(q', y' \mid q, v)$ = probability of entering state q' and emitting output y' given input v and current state q, in system S_i, $v \in U$ or Y_a depending on whether $i=a$ or b.

It is required, of course, that probability be conserved:

$$\sum_{i=1}^{n_{qa}} \sum_{j=1}^{n_{ya}} P_a(q_i^{(a)}, y_j^{(a)} \mid q^{(a)}, u) = 1 \quad \text{for all } q^{(a)} \in Q_a, \quad u \in U$$

$$\sum_{i=1}^{n_{qb}} \sum_{j=1}^{n_{yb}} P_b(q_i^{(b)}, y_j^{(b)} \mid q^{(b)}, y^{(a)}) = 1 \quad \text{for all } q^{(b)} \in Q_b, \quad y^{(a)} \in Y_a$$

and of course, $1 \geqslant P_i(\cdot \mid \cdot) \geqslant 0$ for all arguments $(\cdot)$.

Let Φ_i = set of initial state probabilities associated with system S_i $(i=a,b)$. The overall two-subsystem series may then be succinctly described by the 9-tuple

$$\langle U, Q_a, Q_b, Y_a, Y_b, P_a, P_b, \Phi_a, \Phi_b \rangle$$

where $P=\{P(\cdot \mid \cdot)\}$ = set of conditional probabilities for $i=a,b$, described above.

We now impose the additional constraint that for any finite-length stimulus input, each system yields a finite-length sequence of state changes and

output. Let **0** denote the null output when nothing is emitted from a system and denote an input–state pair that results in **0** by $(v, q)_0$ ($v = u$ or y for S_a or S_b, respectively).

Also required is the "memory" duration of each system, that is, the interval of time until the null output **0** is emitted after the input ceases. Note, though, that **0** might be "emitted" even though inputs are still coming in. In fact, that will be the way of formally describing how a Donderian system appears: No real output emerges until S_a has completed its work on the input.

Let $\mathbf{T}_i(v^*, q^{(i)^*})$ = random duration in S_i from moment of last input v^* and state $q^{(i)^*}$ ($i = a, b$) until no more non-**0** outputs are emitted. If $\mathbf{T}_s$ = input duration as earlier, then we can again decompose the processing times into $\mathbf{T}_s + \mathbf{T}_a(u^*, q^{(a)}) + \mathbf{T}_b(y^*, q^{(b)})$, when u^* = last symbol in the input sequence and y^* = last output from S_a.

A sufficient method of producing a Donderian system is stated in the next, obvious proposition.

Proposition 12.10: Assume S_a and S_b are each eliminatable. If the only output from S_a up until the random time $\mathbf{T}_a$ is a string of **0**s (i.e., no output) then the system is Donderian. □

What could produce such delay of any real output in this type of system? Perhaps in contrast to the typical applications of continuous-time linear systems, many of the tasks of real automata (e.g., computers) are performed by subsystems that emit no output until the task is completed. For instance, when a control processor performs an addition operation on the contents of two memory locations through the accumulator, no output will be apparent (e.g., to an output device or to store in memory) until the addition is finished.

Now consider two experimental factors X_a and X_b (considered as sets) that affect S_a and S_b, respectively, by shortening or prolonging their activity in some way. Most attractive is the possibility that $x_i \in X_i$ might speed up processing in S_i by decreasing the number of operations (i.e., states) through which the system has to move to achieve a certain goal. Alternatively, situations where the durations of state changes might differ from one to another and possibly be altered by experimental manipulation require a general $\Delta t(n, q)$, that is, a variable interval that is a function of input and state. This pulls the theorist out of the bounds of traditionally studied automata. However, one might stay within the constant-Δt systems and still vary *individual* operation times by using the ploy of $q \rightarrow q$ transitions, that is, represent prolonged duration by dwelling in a certain state for the number of Δt required and then transit $q \rightarrow q'$ as appropriate.

It is important to observe that, in general, the shortening or prolongation of $\mathbf{T}_i$ ($i = a, b$) will result in a different train or sequence of states and outputs than the original one. Therefore, because the state and outputs of S_b depend on the outputs of S_a, typically $\mathbf{T}_b$ will be expected to indirectly be a function of X_a as well as X_b.

First, express the factor dependency as $\mathbf{T}_i(v^*, q^{(i)^*}, x_i)$ ($i = a, b$; v^* = last input; $q^{(i)^*}$ = state at moment of last input; $x_i \in X_i$). In the following, it will be postulated that the influences of X_i on S_i ($i = a, b$) occur by virtue of altering the probability distribution P_i, leaving the basic sets Q, Y, unchanged. However, this could readily be generalized without damage to the conclusions below. We content ourselves with a statement of a necessary and sufficient condition for mean RT additivity to hold.

Proposition 12.11: Assume factor X_a does not directly affect S_b in the sense that

$$S_b = \langle Y_a, Q_b, Y_b, P_b, \Phi_b \rangle$$

is unaffected by manipulations of X_a, and similarly for X_b and S_a (recall that Y_a is the input set to S_b). It is then necessary and sufficient for mean RT additivity to hold that

$$\begin{aligned} E_a[E_b(\mathbf{T}_a + \mathbf{T}_b + \mathbf{T}_s)] &= t_s + E_a[\mathbf{T}_a(u^*, q^{(a)^*}, x_a)] \\ &\quad + E_a\{E_b[\mathbf{T}_b(y_a^*, q^{(b)^*}, x_b) \mid x_a]\} \\ &= A(x_a) + B(x_b) \end{aligned}$$

But for this to occur, it is in turn necessary and sufficient that

$$E_a\{E_b[\mathbf{T}_b(y_a^*, q^{(b)^*}, x_b) \mid x_a]\} = B_1(x_a) + B_2(x_b) \tag{12.17}$$

where $E_i(\cdot)$, $i = a, b$, denotes the expectation over the S_a or S_b probability space, respectively.

Proof: Obviously, the first assumption still needs to hold and Eq. 12.17, being the conclusion in Proposition 12.11, is sufficient to produce additivity. Conversely, if Eq. 12.17 does not hold, $\overline{\mathrm{RT}}$ cannot be an additive function of x_a, x_b because Eq. 12.17 would then be a function of x_a also. □

It is not completely transparent what types of probabilistic automata might obey Proposition 12.11 while not actually being Donderian (i.e., nonoverlapping in time). In many cases, the final output of S_a will be different for different x_a, the state of S_b will be different at that time, and the train of states and outputs of S_b will therefore differ as a function of x_a. Another more colloquial way of posing the same conundrum is to ask what kind of system S_b could receive and act in a nontrivial way on inputs from S_a yet not interact with the influences of a factor X_a.

One kind of consideration that might be germane is speed–accuracy tradeoff. Suppose S_a is speeded up (through X_a) without simplifying the task it is doing. Then its output to S_b may be degraded. Now S_b may be able to compensate for its poor input by taking longer (and thus interacting with S_a). Alternatively, it might yield a less accurate output but perform its work during the same time interval (and therefore be affected only by X_b). These

aspects are pertinent even if S_a and S_b are limited in a Donderian fashion. For instance, if $\overline{\text{RT}}$s are additive but error rates increase with faster times, then the RT data by themselves mean little. Further, if the errors increase somewhat faster from perturbations in S_a as well as S_b, it is not clear just what kind of interaction of error rate should be expected; it probably depends on the specific task and subsystems involved. It is situations like those, almost sure to be found in many expected settings, that beseech the use of substantive mathematical information-processing models in order to test specific theoretical predictions and to use the parameters to measure important processing characteristics.

Summary and conclusions

In attempting to summarize the rather diverse results of this chapter we make the following observations:

1. An additive factor approach is here viewed as important in helping to delineate the type of interactions that various subsystems may have with one another. In particular, mean RT additivity is possible with serially arranged subsystems with no overlap in processing times, whereas mean RT interactions are compatible with certain broad classes of parallel systems. Heretofore, no broad classes of models failing to exhibit mean RT additivity had been demonstrated. However, questions regarding the statistical power of the usual mean RT additivity tasks have not been probed.

2. It was argued that the great majority of RT theorists implicitly assume that RT lies on a ratio scale but that this assumption is, at present, untestable. The use of conjoint measurement to test for RT additivity was considered, and it was concluded that conjoint measurement offers few advantages over more traditional techniques such as analysis of variance. Dimensional analysis was briefly considered with respect to its potential use in behavioral psychology and, in particular, RT. Although a powerful tool in physics and engineering, it is not presently suitable for application in the nonphysical sciences.

3. Sequentially arranged linear systems and sequential finite probabilistic automata were considered as examples of systems in which processing times generally overlap, thus producing hybrid or, in special cases, parallel or serial systems. Formulations that result in serial (nonoverlapping) times or mean RT additivity were developed. However, more work needs to be done with regard to actual information-processing operations that might occur in such systems that obey these formulations.

Appendix 12.1

To derive Eq. 12.15 in the discrete case note that the discrete analogy of Eq. 12.14 is

$$y(k\Delta t)=\sum_{j=0}^{k} h_b(k\Delta t-j\Delta t)\sum_{i=0}^{j} h_a(j\Delta t-i\Delta t)u(i\Delta t) \tag{A.1}$$

Now let us suppress the Δt in order to shorten our notation, but we always keep in mind that, for instance, j refers to $j\Delta t$ time units. Now we write Eq. A.1 in detail:

$$\begin{aligned}y(k\Delta t)=&u(0)h_a(0)h_b(k)+u(0)h_a(1)h_b(k-1)+u(1)h_a(0)h_b(k-1)\\&+u(0)h_a(2)h_b(k-2)+u(1)h_a(1)h_b(k-2)+u(2)h_a(0)h_b(k-2)\\&+\cdots+u(0)h_a(k)h_b(0)+u(1)h_a(k-1)h_b(0)\\&+\cdots+u(k)h_a(0)h_b(0)\end{aligned}$$

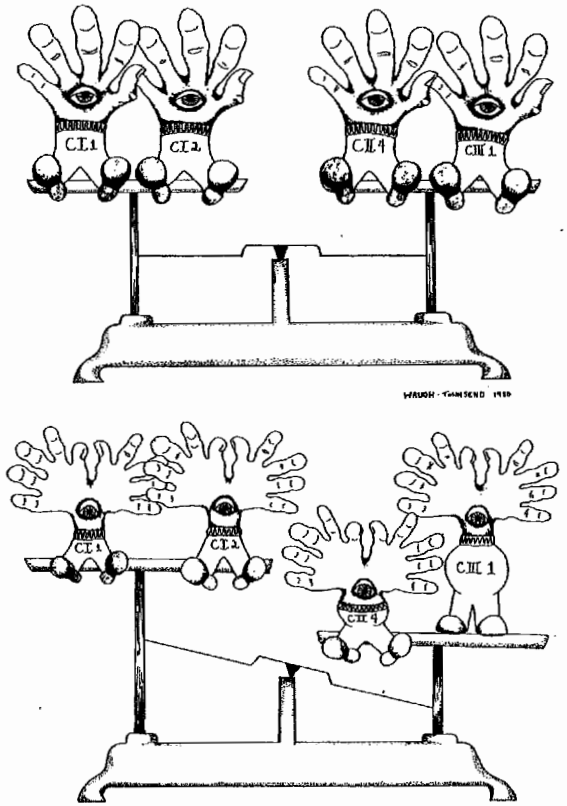
A little scrutiny of this sum suggests an alternative method of forming the identical thing:

$$\begin{aligned}y(k\Delta t)=&u(0)[h_a(0)h_b(k)+h_a(1)h_b(k-1)+h_a(2)h_b(k-2)\\&+\cdots+h_a(k)h_b(0)]\\&+u(1)[h_a(0)h_b(k-1)+h_a(1)h_b(k-2)+\cdots+h_a(k-1)h_b(0)]\\&+u(2)[h_a(0)h_b(k-2)+h_a(1)h_b(k-3)\\&+\cdots+h_a(k-2)h_b(0)]\\&+\cdots+u(k-1)[h_a(0)h_b(1)+h_a(1)h_b(0)]+u(k)h_a(0)h_b(0)\end{aligned}$$

We can summarize this result with our standard summation signs as

$$y(k\Delta t)=\sum_{j=0}^{k} u(j)\sum_{i=0}^{k-j} h_a(i)h_b(k-j-i)$$

which is the discrete analog of Eq. 12.15.



The Handpeople reveal two contrasting predictions of *parallel vs. serial models* in an experimental design especially constructed to test between reasonably large classes of parallel and serial systems. Each of the Handpersons stands for one of the critical experimental conditions (for this particular test). It can be seen that the serial models (*top*) predict that the sum of conditions CI 1 and CI 2 must equal the sum of CII 4 and CIII 1. On the other hand, the parallel models (*bottom*) predict that this sum will be unequal. Both parallel and serial models assume the ability to self-terminate (see also Chapter 4).

13 *The parallel–serial testing paradigm*

An experimental paradigm has been developed that capitalizes on differentiating characteristics of parallel and serial processes (Townsend 1976a; also see Chapters 4, 14, and 15). For some broad classes of parallel and serial models, it can be shown that, for the conditions that comprise the paradigm, no member of either class can make exactly the same predictions as the other. Its logical acronym is PST (parallel–serial tester). It will be defined precisely below. One of the nice properties of the paradigm is that the testing can be at the level of the mean reaction times. This obviates dilemmas with the exact form of the distribution (being forced to posit a particular distribution, concern about the large sampling variance of the higher sample moments, problems of estimating the RT density function, and so on). On the other hand, when it is desired to employ specific classes of distributions and to test the

parallel vs. serial predictions at the distributional level, it may be that the sensitivity of the test will be enhanced.

Another approach that makes no specific assumption about the underlying distribution tests the tails of the observable RT latency density in order to ascertain whether it could have arisen from a within-stage independent parallel process (Townsend 1976b). Ross and Anderson (1981) have adapted this method for use with convolutions of mixtures of distributions and employed it with data to test an implication of Anderson's ACT theory (Anderson 1976).

The plan of the chapter will be to begin with an introduction to the basic empirical paradigm itself followed by a section describing the parallel and serial models' notation. Then, the models' predictions within PST are presented and three propositions concerning parallel–serial testability are put forth and proven. The propositions given here are more general than those of Townsend (1976a) and generally follow the form of more recent work (Townsend 1976b). Both the serial and parallel models covered are distribution-free. The succeeding section gives a special case and some numerical examples. An experimental application of PST carried out by Townsend and Snodgrass (1974) is then explored, with evidence for parallel processing in that particular instance. The final portion of the chapter briefly considers questions related to generalizability of these results.

The basic paradigm

The PST paradigm requires an observer to search through a list of two items for one or more targets. In principle, it is irrelevant whether the target is presented first, second, or simultaneously with the two comparison items. That is, the focus is on comparison times rather than encoding and other attendant processes.

For convenience we will refer to the target as A, although it may be any member of the stimulus set. Then the paradigm is constituted by the three experimental conditions listed in Table 13.1. Condition CI is composed of two types of trials, and CII and CIII of four types of trials each, where the trials are assumed to be randomized within conditions and the three conditions are blocked separately. One of two responses is required on every trial, and these are designated R1 and R2.[1]

Condition I requires response R1 if the target item is on the right (spatially, temporally, etc.) and R2 if the target is on the left. Condition CII, on the other hand, requires that both comparison items match the target in order for response R1 to be appropriate, whereas in CIII only one comparison is required to match the target in order for R1 to be the correct response. At

[1] The response pairing with the trial types is slightly different from that in Townsend (1976a) in order to be compatible with the empirical application in a later section of this chapter.

Table 13.1. *Parallel-serial testing (PST) paradigm*

Condition			
CI	Target = A	Comparison items	Response
	Trial type	1. AB	R2
		2. BA	R1
CII	Target = A	Comparison items	Response
	Trial type	1. AA	R1
		2. AB	R2
		3. BA	R2
		4. BB	R2
CIII	Target = A	Comparison items	Response
	Trial type	1. AA	R1
		2. AB	R1
		3. BA	R1
		4. BB	R2

present, the theorems concerning PST require that errors be irrelevant to the basic comparison process. Although there is rightfully a good deal of interest in speed–accuracy relations (and a significant degree of attention on it in this treatment; see the Preface and Chapters 5, 6, 9–11, and 15), as mentioned earlier, it is not yet clear just what role, if any, quite small error rates (e.g., less than 10%) play in perceptual matching. It even seems doubtful that RT–error covariation would be likely to cause a parallel PST outcome to appear serial or vice versa, but such questions deserve further quantitative investigation. Krueger (1978) contributes an intriguing theoretical analysis of RT and error data in a class of matching experiments similar to these.

The logical structure of the PST is important rather than the particular realizations. Note that in the prototypical characterization above, a single target (A) is shown. In the application reported below, two symbols actually comprise the target rather than simply one. Similarly, the stimuli presented may, of course, vary from trial to trial, and the "matching" could be in terms of categories (or anything) rather than physical identity.

The basic models

Two postulates are shared by all the models. One is that processing is basically self-terminating and the other is that the distribution on processing time is somehow different when two matching items are being compared than it is when two mismatching items are compared. In addition, it is assumed that the extra, residual processing times that come from outside the comparison process are invariant across the various conditions of the paradigm, although they can be assumed to be different in the parallel and serial models without

disabling the testability results. The comparison and residual processing times are also assumed to be additive but need not be stochastically independent because in the present treatment it is the mean reaction times that are utilized. As we will see, under these conditions PST yields observable diversity (i.e., distinguishability) at the level of the mean (Townsend 1976a). An experimental condition produces mean-observable diversity if the mean RTs are observable (trivial in most settings) and if the two models make distinct mean RT predictions. To economize the terminology we shall again (e.g., as in Chapter 11) refer to the general stochastic processes appropriate for a situation as the model, although in a strict sense it may be a class of models (because it generates a family of particular probability distributions on RTs).

Self-termination, meaning the ability of the system to stop processing when sufficient information has been obtained to respond correctly, certainly occurs in some cognitive tasks. For instance, when searching for the name of an old acquaintance, one need not exhaust the entire long-term memory reservoir after the correct (or incorrect) one has been retrieved. Its existence in other contexts, such as short-term memory search, is more controversial. Serial position effects are often cited as evidence for self-termination, and we find these fairly compelling (see Townsend 1974b), but there are other ways to predict them without assuming self-termination. The reader is referred to Chapters 6 and 7 for more discussion of this and similar points.

The idea that it might take a different amount of time to complete a comparison of matching elements than it does to compare mismatching elements, or more accurately, that matching and nonmatching elements possess a different distribution on completion time, has also received a good deal of support in the perceptual and cognitive literature. Examples can be found in same–different matching (Bamber 1969, 1975; Taylor 1976b), psycholinguistics (Chase & Clark 1972), and memory or display search (Townsend & Roos 1973; Huesman & Woocher 1976). In many cases, same-match (+) RTs appear to be faster than different-match (−) RTs, but in especially simple psychophysical circumstances (e.g., with unidimensional stimuli; see Krueger 1978) − matches may be faster than + matches.

The additive decomposition of the mean times into the process of interest plus all the others lumped together in a residual latency term is not without criticism (see the non-Donderian model of Chapter 6; see also Chapter 12), but seems a reasonable place to begin.

An important postulate with regard to PST is that the primary characteristics of processing do not change across conditions CI, CII, and CIII. This requires that the aforementioned postulates hold in all conditions and in addition that the basic parameter set be invariant across conditions. Here we face a dilemma associated with virtually all scientific theory testing: The broader the class of models tested, the more complex the experimental conditions must be to permit testability, but the more complex the experimental conditions, the less likely it is that models of the same subsystems operating with the same parameter values are applicable across all experimental condi-

tions. All we can do in such circumstances is attempt to maximize the likelihood that the same basic model parameters are indeed appropriate everywhere that is required. Now assume that a is the left position in Table 13.1 and b the right position.

Let $\mathbf{T}_{ij}^{+}$ be the random time to compare the item in position i $(i=a,b)$ with the target A when the comparison is taking place during stage j $(j=1,2)$, and the superscript + indicate that the two items match; $\mathbf{T}_{ij}^{-}$ is the same duration when the items mismatch. Note that the **T** terms are intercompletion times rather than completion times (see Chapters 3 and 4). The models below do not assume across-stage independence (i.e., independence of intercompletion times; see Chapters 3 and 4). However, special exponentially based models developed later do.

The serial models

We can simply use the mean times $E(\mathbf{T}_{ij}^{\bullet})=\bar{t}_{ij}^{\bullet}$, where $\bullet$ stands for + or −, in the proof below, because the mean of the sum of the serial times is, of course, just the sum of the respective means. As usual, p is the probability that the item in position a is selected first for processing. The later part of the chapter requires the distributions, however, so we establish the requisite notation here.

Let $F_{i1}(t_{i1}^{\bullet})$ be the distribution function for the comparison of position i's element with the target during stage 1, where again $\bullet$ may be either + or −, and $F_{j2}(t_{j2}^{\bullet}\,|\,t_{i1}^{\bullet})$ be the distribution function during stage 2 on position j, conditioned on the duration occupied by the comparison of position i in stage 1. These functions are, in general, not constrained to be even from the same family of distributions, although that might typically be the case. As customary, $f=dF/dt$ is the appropriate density function.

The parallel models

Recall first that in our nomenclature *stage* refers to the interval between two successive item (element) completions. Thus, stage 1 designates the interval before the first comparison finishes, stage 2 that between the first and second completions, and so on. An alternate definition of *stage* is as one of a series of subprocesses (see Chapter 12).

It will be assumed that the parallel models are within-stage-independent (see Townsend 1976a and Chapter 4). As usual, let the survivor function of the intercompletion time be $\bar{G}(t)=1-G(t)$ (i.e., 1 − distribution function). Then within-stage independence says that the survivor function of the intercompletion time is expressible as a product of the survivor functions of the uncompleted elements (see Chapter 4):

$$P(\mathbf{T}_{a1}>t\cap\mathbf{T}_{b1}>t)=\bar{G}_{a1}(t)\bar{G}_{b1}(t)$$

where $\bar{G}_{a1}(t)$ and $\bar{G}_{b1}(t)$ are the survivor functions of the two positions during stage 1 and it is not necessary that $\bar{G}_{a1}(t)=\bar{G}_{b1}(t)$ even when both matches are positive or negative [e.g., $\bar{G}_{a1}^{+}(t)\neq\bar{G}_{b1}^{+}(t)$]. (Note that the bar over G is not to be interpreted as a minus sign.) There is only one uncompleted element during stage 2, of course, and its survivor function is $\bar{G}_{i2}(t_{i2})$. Finally, $g=-d\bar{G}/dt=dG/dt$ is the parallel density of the distribution function $G=1-\bar{G}$, and we are obviously assuming the G terms are differentiable functions on $t\in[0,\infty)$. It is also assumed that the $\bar{G}$ terms are strictly decreasing in t. Doubtless, it would be possible to weaken these last restrictions somewhat, but there are cases where parallel models can mimic serial models by assuming, say,

$$\begin{aligned}\bar{G}_{ij}(t)&=1, \quad t=0\\ &=0, \quad t>0\end{aligned}\qquad\text{(element is processed with infinite speed)}$$

for certain of the $\bar{G}_{ij}$ terms. In such cases, the parallel model degenerates into a serial model. These circumstances are not permitted in the propositions, however, in view of the assumed differentiability and strict monotonicity of the $\bar{G}$ terms.

Let $I(A)$ be an indicator random variable on the event A such that

$$\begin{aligned}I(A)&=1 \quad \text{if } A \text{ occurs}\\ &=0 \quad \text{otherwise}\end{aligned}$$

Consider $A=\mathbf{T}_{a1}^{+}<\mathbf{T}_{b1}^{-}$, that is, b is completed after a. Then it will be convenient to use joint expectations of the form

$$E[\mathbf{T}_{b2}^{-}I(A)]=E[\mathbf{T}_{b2}^{-}I(\mathbf{T}_{a1}^{+}<\mathbf{T}_{b1}^{-})]$$

which yields the expected second-stage intercompletion time on b when it is a mismatch and occurs after a's completion, which was a positive match. The conditional expectation is, of course, easily computed from this to be

$$E(\mathbf{T}_{b2}^{-}\,|\,\mathbf{T}_{a1}^{+}<\mathbf{T}_{b1}^{-})=\frac{E[\mathbf{T}_{b2}^{-}I(\mathbf{T}_{a1}^{+}<\mathbf{T}_{b1}^{-})]}{P(\mathbf{T}_{a1}^{+}<\mathbf{T}_{b1}^{-})}$$

Predictions and propositions

The mean reaction time predictions for both kinds of models are shown in Table 13.2. The serial predictions are straightforward. Thus, the prediction for CII 1 is derived by noting that both comparisons always have to be made and with probability p position a is selected first ($\bar{t}_{a1}^{+}$) followed by position b ($\bar{t}_{b2}^{+}$), both matches are +, and so on. The $\bar{t}_s$ equals the combined residual durations from extraneous processes.

The logic of the parallel derivations with regard to when processing can terminate is similar but slightly more complex. We break the contributions down into that for the minimum (first completion) time plus the second intercompletion time. For instance, consider the nongestalt parallel expression for condition CII 1. Because both comparisons must be finished in order to

Table 13.2. *PST Predictions*

	Serial models		Parallel models
CI	1. $p\bar{t}_{a1}^{+}+(1-p)\bar{t}_{b1}^{-}+\bar{t}_s$	CI	1. $\int_0^\infty \bar{G}_{a1}^{+}(t)\,\bar{G}_{b1}^{-}(t)\,dt+\bar{t}_p$
	2. $p\bar{t}_{a1}^{-}+(1-p)\bar{t}_{b1}^{+}+\bar{t}_s$		2. $\int_0^\infty \bar{G}_{a1}^{-}(t)\,\bar{G}_{b1}^{+}(t)\,dt+\bar{t}_p$
CII	1. $p(\bar{t}_{a1}^{+}+\bar{t}_{b2}^{+})+(1-p)(\bar{t}_{b1}^{+}+\bar{t}_{a2}^{+})+\bar{t}_s$	CII	1. $\int_0^\infty \bar{G}_{a1}^{+}(t)\,\bar{G}_{b1}^{+}(t)\,dt+E[\mathbf{T}_{b2}^{+}I(\mathbf{T}_{a1}^{+}<\mathbf{T}_{b1}^{+})]+E[\mathbf{T}_{a2}^{+}I(\mathbf{T}_{b1}^{+}<\mathbf{T}_{a1}^{+})]+\bar{t}_p$
			Gestalt prediction: $E(\mathbf{T}_{ab}^{++})+\bar{t}_p$
	2. $p(\bar{t}_{a1}^{+}+\bar{t}_{b2}^{-})+(1-p)\bar{t}_{b1}^{-}+\bar{t}_s$		2. $\int_0^\infty \bar{G}_{a1}^{+}(t)\,\bar{G}_{b1}^{-}(t)\,dt+E[\mathbf{T}_{b2}^{-}I(\mathbf{T}_{a1}^{+}<\mathbf{T}_{b1}^{-})]+\bar{t}_p$
	3. $p\bar{t}_{a1}^{-}+(1-p)(\bar{t}_{b1}^{+}+\bar{t}_{a2}^{-})+\bar{t}_s$		3. $\int_0^\infty \bar{G}_{a1}^{-}(t)\,\bar{G}_{b1}^{+}(t)\,dt+E[\mathbf{T}_{a2}^{-}I(\mathbf{T}_{b1}^{+}<\mathbf{T}_{a1}^{-})]+\bar{t}_p$
	4. $p\bar{t}_{a1}^{-}+(1-p)\bar{t}_{b1}^{-}+\bar{t}_s$		4. $\int_0^\infty \bar{G}_{a1}^{-}(t)\,\bar{G}_{b1}^{-}(t)\,dt+\bar{t}_p$
CIII	1. $p\bar{t}_{a1}^{+}+(1-p)\bar{t}_{b1}^{+}+\bar{t}_s$	CIII	1. $\int_0^\infty \bar{G}_{a1}^{+}(t)\,\bar{G}_{b1}^{+}(t)\,dt+\bar{t}_p$
	2. $p\bar{t}_{a1}^{+}+(1-p)(\bar{t}_{b1}^{-}+\bar{t}_{a2}^{+})+\bar{t}_s$		
	3. $p(\bar{t}_{a1}^{-}+\bar{t}_{b2}^{+})+(1-p)\bar{t}_{b1}^{+}+\bar{t}_s$		Gestalt prediction: $E(\mathbf{T}_{ab}^{++})+\bar{t}_p$
	4. $p(\bar{t}_{a1}^{-}+\bar{t}_{b2}^{-})+(1-p)(\bar{t}_{b1}^{-}+\bar{t}_{a2}^{-})+\bar{t}_s$		2. $\int_0^\infty \bar{G}_{a1}^{+}(t)\,\bar{G}_{b1}^{-}(t)\,dt+E[\mathbf{T}_{a2}^{+}I(\mathbf{T}_{b1}^{-}<\mathbf{T}_{a1}^{+})]+\bar{t}_p$
			3. $\int_0^\infty \bar{G}_{a1}^{-}(t)\,\bar{G}_{b1}^{+}(t)\,dt+E[\mathbf{T}_{b2}^{+}I(\mathbf{T}_{a1}^{-}<\mathbf{T}_{b1}^{+})]+\bar{t}_p$
			4. $\int_0^\infty \bar{G}_{a1}^{-}(t)\,\bar{G}_{b1}^{-}(t)\,dt+E[\mathbf{T}_{b2}^{-}I(\mathbf{T}_{a1}^{-}<\mathbf{T}_{b1}^{-})]+E[\mathbf{T}_{a2}^{-}I(\mathbf{T}_{b1}^{-}<\mathbf{T}_{a1}^{-})]+\bar{t}_p$

insure a correct response, exhaustive processing is required. Thus, the mean RT is equal to the sum $\mathbf{T}_1+\mathbf{T}_2$, where $\mathbf{T}_i$ is the ith intercompletion time ($i=1,2$), plus $\bar{t}_p$, the mean residual processing time. The first-stage duration $\mathbf{T}_1^{+}$ is just $\min(\mathbf{T}_{a1}^{+},\mathbf{T}_{b1}^{+})$, whereas the second, $\mathbf{T}_2$, is $\mathbf{T}_{a2}^{+}$ or $\mathbf{T}_{b2}^{+}$ depending on whether b or a was finished first. Therefore,

$$E(\mathbf{T}_1+\mathbf{T}_2)=E[\min(\mathbf{T}_{a1}^{+},\mathbf{T}_{b1}^{+})]\quad\text{stage 1}$$
$$+E[\mathbf{T}_{b2}^{+}I(\mathbf{T}_{a1}^{+}<\mathbf{T}_{b1}^{+})]+E[\mathbf{T}_{a2}^{+}I(\mathbf{T}_{b1}^{+}<\mathbf{T}_{a1}^{+})]\quad\begin{matrix}\text{stage 2 and } a \text{ is first}\\ \text{stage 2 and } b \text{ is first}\end{matrix}$$

But we know that the expected duration is the integral of the survivor function (Chapter 3), which, for the minimum duration here, is just

$$P[\min(\mathbf{T}_{a1}^{+},\mathbf{T}_{b1}^{+})>t]=\bar{G}_{a1}^{+}(t)\,\bar{G}_{b1}^{+}(t)$$

so that

$$E(\mathbf{T}_1+\mathbf{T}_2)=\int_0^\infty \bar{G}_{a1}^{+}(t)\,\bar{G}_{b1}^{+}(t)\,dt+E[\mathbf{T}_{b2}^{+}I(\mathbf{T}_{a1}^{+}<\mathbf{T}_{b1}^{+})]+E[\mathbf{T}_{a2}^{+}I(\mathbf{T}_{b1}^{+}<\mathbf{T}_{a1}^{+})]$$

The latter terms can be explicated mathematically as in

$$E[\mathbf{T}_{b2}^{+}I(\mathbf{T}_{a1}^{+}<\mathbf{T}_{b1}^{+})]=\int_0^\infty\int_0^\infty t_{b2}^{+}g_{a1}^{+}(t_{a1}^{+})\,\bar{G}_{b1}^{+}(t_{a1}^{+})\,g_{b2}^{+}(t_{b2}^{+}\mid t_{a1}^{+})\,dt_{a1}^{+}\,dt_{b2}^{+}$$

Of course, in a number of the conditions only the minimum time is pertinent (CI 1, 2; CII 4; CIII 1) or only the minimum time plus that concerned with a particular order of processing (CII 2, 3; CIII 2, 3). However, the derivations are similar.

Some early pilot work indicated that when both elements matched the target, as in CII 1 and CIII 1, processing seemed to be quite fast and about equal on the two conditions, although the standard serial or non-supercapacity parallel prediction is that CIII 1 < CII 1. This is because termination can occur in CIII 1 as soon as the first element is completed, whereas both must always be finished in CII 1. This suggests that observers were able to treat the double match situation as a single global gestalt pattern. For this reason, a separate gestalt prediction is included in the parallel column for CII 1 and CIII 1. If the overall comparison pattern is conceived of as a gestalt, we can simply write the expectation for that stimulus. These models will be fit to data separately below.

Propositions 13.1 and 13.2 are concerned with the regular parallel model. Following a brief discussion of those, Proposition 13.3 treats the gestalt case.

The first proposition claims that if $\bar{G}_{a1}^{+}(t)=\bar{G}_{b1}^{+}(t)=\bar{G}_1^{+}(t)$ and if $\bar{G}_{a1}^{-}(t)=\bar{G}_{b1}^{-}(t)=\bar{G}_1^{-}(t)$ for all $t>0$ but that

$$E(\mathbf{T}_1^{-})=\int_0^\infty \bar{G}_1^{-}(t)\,dt\neq\int_0^\infty \bar{G}_1^{+}(t)\,dt=E(\mathbf{T}_1^{+})$$

then the PST predictions can never be equal for the parallel and serial models. It is therefore assumed that no serial position effects exist during stage 1 ($\bar{G}^{\cdot}_{a1}=\bar{G}^{\cdot}_{b1}$; $\cdot$ equals $+$, $-$) but in addition we only assume that the $+$ and $-$ mean times differ. This latter condition is not generally strong enough when serial position effects are present during stage 1 to prove the result because of an excess of flexibility in the models. Proposition 13.2 takes up this case and it is shown that if the $+$ and $-$ distributions are ordered, that is, if

$$\bar{G}^{+}_{a1}(t)\gtrless\bar{G}^{-}_{a1}(t) \quad \text{and} \quad \bar{G}^{+}_{b1}(t)\gtrless\bar{G}^{-}_{b1}(t) \quad \text{for all } t\gtrless 0$$

then the parallel and serial predictions are distinct. This concept of an ordering of the distributions is one reasonable but relatively strong way two probability distributions can differ. For instance, a weaker condition is simply that the means are ordered, as in Proposition 13.1, whereas a stronger condition is that the hazard functions are ordered for all $t\gtrless 0$. The reader is referred to Chapters 8 and 9 and to Townsend and Ashby (1978) for more discussion of these relationships.

In the first two propositions, parallel–serial discriminability also occurs if $\bar{t}^{+}_{a1}\neq\bar{t}^{-}_{a1}$ and $\bar{t}^{+}_{b1}\neq\bar{t}^{-}_{b1}$ in the serial model, if it is supposed that when the parallel distributions differ, they do so in the manner just described. Actually, the following more general result can be proven: If the $+$ and $-$ difference only exists at one serial position (say, a), then enough significant constraints are generated in the classes of models to make them PST observable mean diverse (as in the more special models of Townsend 1976a). That is, they can still be experimentally discriminated at the level of mean RTs. However, the present treatment is sufficient to illustrate the main points of this type of analysis.

The main thrust of Propositions 13.1 and 13.2 is associated with property IIC of Chapter 15, one of the differentiating aspects of parallel and serial functioning. That property maintains that the processing time of an element operated on serially can not reasonably depend on the identity of elements completed at a later time, whereas such a dependence is reasonable when processing is parallel. This idea is implemented by adding conditions CI 1 + CI 2 and comparing the result with CII 4 + CIII 1. In the serial model property IIC constrains both sums to be an equal and additively separable function of the $+$ and $-$ processing durations; that is, CI 1 + CI 2 = CII 4 + CIII 1 because each intercompletion time is forced to be a single element's processing time (and independent of what transpires later). In the parallel model, however, the intercompletion times are products of the $+$ and $-$ distributions and cannot be so decomposed. Because the sums CI 1 + CI 2 and CII 4 + CIII 1 are distinctive in the parallel instance, CI 1 + CI 2 $\neq$ CII 4 + CIII 1.[2]

The fact that a single diverse property of realistic parallel and serial systems, together with help from the structure of PST, has led to a powerful dis-

[2] D. Bamber (personal communication) has noted that CI 1 + CI 2 = CII 3 + CIII 1 even if $\bar{t}_s$ varies from response R1 to R2.

criminatory experimental paradigm may inculcate some optimism in us regarding the future of parallel–serial testability, as well as other tricky psychological issues.

Proposition 13.1: Parallel and serial models are observable mean diverse in PST if

(i) $\bar{G}^{+}_{1}(t)=\bar{G}^{+}_{a1}(t)=\bar{G}^{+}_{b1}(t)$ and $\bar{G}^{-}_{1}(t)=\bar{G}^{-}_{a1}(t)=\bar{G}^{-}_{b1}(t)$ for all $t>0$ and $\bar{G}^{+}_{1}(t)\neq\bar{G}^{-}_{1}(t)$ for some $t>0$ in the parallel model; or

(ii) $\bar{t}^{+}_{a1}\neq\bar{t}^{-}_{a1}$ and $\bar{t}^{+}_{b1}\neq\bar{t}^{-}_{b1}$ in the serial model.

Proof: (i) From Table 13.2 observe that if mean RTs are everywhere the same; then under the above conditions,

$$\begin{aligned}\text{CI 1}+\text{CI 2}&=p(\bar{t}^{+}_{a1}+\bar{t}^{-}_{a1})+(1-p)(\bar{t}^{+}_{b1}+\bar{t}^{-}_{b1})+2\bar{t}_s\\&=\int_0^\infty \bar{G}^{+}_{1}(t)\bar{G}^{-}_{1}(t)\,dt+\int_0^\infty \bar{G}^{-}_{1}(t)\bar{G}^{+}_{1}(t)\,dt+2\bar{t}_p\\&=2\int_0^\infty \bar{G}^{+}_{1}(t)\bar{G}^{-}_{1}(t)\,dt+2\bar{t}_p\end{aligned}$$

But also

$$\begin{aligned}\text{CII 4}+\text{CIII 1}&=p(\bar{t}^{+}_{a1}+\bar{t}^{-}_{a1})+(1-p)(\bar{t}^{+}_{b1}+\bar{t}^{-}_{b1})+2\bar{t}_s\\&=\int_0^\infty [\bar{G}^{-}_{1}(t)]^2\,dt+\int_0^\infty [\bar{G}^{+}_{1}(t)]^2\,dt+2\bar{t}_p\end{aligned}$$

so the implication is that CI 1 + CI 2 = CII 4 + CIII 1 and hence

$$\int_0^\infty [\bar{G}^{-}_{1}(t)]^2\,dt+\int_0^\infty [\bar{G}^{+}_{1}(t)]^2\,dt=2\int_0^\infty \bar{G}^{+}_{1}(t)\bar{G}^{-}_{1}(t)\,dt$$

which gives

$$\int_0^\infty [\bar{G}^{-}_{1}(t)-\bar{G}^{+}_{1}(t)]^2\,dt=0$$

We need to show this cannot be.

It was supposed that at least one t, say t^*, existed where the two functions differed. It is straightforward to show that there must be an interval around such a t^* where the functions still differ, because of the assumed continuity of the G functions. Presume that at t^*,

$$|\bar{G}^{+}(t^*)-\bar{G}^{-}(t^*)|=h>0$$

Now, by continuity, for any $\epsilon_1,\epsilon_2>0$, there exists a $\delta>0$ such that if $|t-t^*|<\delta$, then

$$|\bar{G}^{+}(t)-\bar{G}^{+}(t^*)|<\epsilon_1 \quad \text{and} \quad |\bar{G}^{-}(t)-\bar{G}^{-}(t^*)|<\epsilon_2$$

Therefore, for any t such that $|t-t^*|<\delta$,

$$|\bar{G}^+(t)-\bar{G}^-(t)|$$
$$=|\bar{G}^+(t^*)-\bar{G}^-(t^*)+\bar{G}^+(t)-\bar{G}^+(t^*)-[\bar{G}^-(t)-\bar{G}^-(t^*)]|$$
$$\geqslant h-|\epsilon_1+\epsilon_2|$$

But because ϵ_1 and ϵ_2 are arbitrarily small, it is clear that an interval around t^* exists where $\bar{G}_1^+ \neq \bar{G}_1^-$. This implies that

$$\int_0^\infty [\bar{G}_1^+(t)-\bar{G}_1^-(t)]^2\,dt>0$$

ending in a contradiction.

(ii) Suppose $\bar{t}_{a1}^+ \neq \bar{t}_{a1}^-$ and $\bar{t}_{b1}^+ \neq \bar{t}_{b1}^-$ and notice that then either

$$\text{CI 1}=p\bar{t}_{a1}^+ + (1-p)\bar{t}_{b1}^- \neq p\bar{t}_{a1}^- + (1-p)\bar{t}_{b1}^- = \text{CII 4}$$

or

$$\text{CI 2}=p\bar{t}_{a1}^- + (1-p)\bar{t}_{b1}^+ \neq p\bar{t}_{a1}^- + (1-p)\bar{t}_{b1}^- = \text{CII 4}$$

irrespective of the value of p. Therefore, in the parallel model,

$$\int_0^\infty \bar{G}_1^+(t)\bar{G}_1^-(t)\,dt \neq \int_0^\infty [\bar{G}_1^-(t)]^2\,dt$$

This implies that $\bar{G}_1^+(t) \neq \bar{G}_1^-(t)$ for some value of t, and again the result comes about through the foregoing argument. □

Proposition 13.1 holds no matter what the value of the serial parameter p; specifically, it need not be constrained to the open interval $(0, 1)$. If it is (i.e., if $p \neq 0, 1$), then the condition on the mean serial times can be weakened to $\bar{t}_{a1}^+ \neq \bar{t}_{a1}^-$ *or* $\bar{t}_{b1}^+ \neq \bar{t}_{b1}^-$.

Part (i) of Proposition 13.2 states a sufficient condition for observable mean diversity that is in the form of a condition placed directly on the parallel functions. Part (ii) supposes that the serial $\bar{t}_{a1}^+ \neq \bar{t}_{a1}^-$, or $\bar{t}_{b1}^+ \neq \bar{t}_{b1}^-$. Then under the assumption that *if* any parallel difference is present, it appears in the ordering on survivor functions $\bar{G}_{a1}^+(t) \gtrless \bar{G}_{a1}^-(t)$ or $\bar{G}_{b1}^+(t) \gtrless \bar{G}_{b1}^-(t)$, mean observable diversity again ensues.

Proposition 13.2: Mean diversity holds in PST,

(i) if for all $t>0$, $\bar{G}_{a1}^+(t) \neq \bar{G}_{a1}^-(t)$ and $\bar{G}_{b1}^+(t) \neq \bar{G}_{b1}^-(t)$, or

(ii) if $\bar{t}_{a1}^+ \neq \bar{t}_{a1}^-$, $\bar{t}_{b1}^+ \neq \bar{t}_{b1}^-$ and the only way that the $\bar{G}$ terms are allowed to differ is by being distinctive at all values of t (which implies and is implied by an ordering on the survivor functions $\bar{G}_{\cdot 1}^+ > \bar{G}_{\cdot 1}^-$ or vice versa).

Proof: (i) Again, scrutiny of the serial side of Table 13.2 yields CI 1 + CI 2 = CII 4 + CIII 1, which implies the parallel constraint

$$\int_0^\infty [\bar{G}_{a1}^+(t)-\bar{G}_{a1}^-(t)][\bar{G}_{b1}^-(t)-\bar{G}_{b1}^+(t)]\,dt=0$$

By hypothesis, $\bar{G}_{a1}^+ \neq \bar{G}_{a1}^-$ and $\bar{G}_{b1}^+ \neq \bar{G}_{b1}^-$ for all $t>0$, so they can never cross; therefore,

$$\bar{G}_{a1}^+(t) \gtrless \bar{G}_{a1}^-(t) \quad \text{and} \quad \bar{G}_{b1}^-(t) \gtrless \bar{G}_{b1}^+(t) \quad \text{for all } t>0$$

so the left-hand side cannot integrate to 0 in the parallel constraint and the result is gained by contradiction.

(ii) If $\bar{t}_{a1}^+ \neq \bar{t}_{a1}^-$ and $\bar{t}_{b1}^+ \neq \bar{t}_{b1}^-$, as in Proposition 13.1, we find that CI 1 ≠ CII 4 and CI 2 ≠ CII 4 so that

$$\int_0^\infty \bar{G}_{a1}^+(t)\bar{G}_{b1}^-(t)\,dt \neq \int_0^\infty \bar{G}_{a1}^-(t)\bar{G}_{b1}^-(t)\,dt$$

and

$$\int_0^\infty \bar{G}_{a1}^-(t)\bar{G}_{b1}^+(t)\,dt \neq \int_0^\infty \bar{G}_{a1}^-(t)\bar{G}_{b1}^-(t)\,dt$$

Then, by assumption, $\bar{G}_{a1}^+(t) \neq \bar{G}_{a1}^-(t)$ and $\bar{G}_{b1}^+(t) \neq \bar{G}_{b1}^-(t)$ for all $t>0$ and the result follows. □

Observe that part (i) of Proposition 13.2 did not require the distribution ordering postulate of part (ii) because it was implied by $\bar{G}_{i1}^+(t) \neq \bar{G}_{i1}^-(t)$ $(i=a, b)$ for *any* $t>0$.

As it turns out, in most empirical applications the predictions of parallel and serial models will differ in PST if

$$E(\mathbf{T}_{a1}^+)=\int_0^\infty \bar{G}_{a1}^+(t)\,dt \neq \int_0^\infty \bar{G}_{a1}^-(t)\,dt=E(\mathbf{T}_{a1}^-)$$

and

$$E(\mathbf{T}_{b1}^+)=\int_0^\infty \bar{G}_{b1}^+(t)\,dt \neq \int_0^\infty \bar{G}_{b1}^-(t)\,dt=E(\mathbf{T}_{b1}^-)$$

It is usually not necessary that $\bar{G}_{a1}^+(t) \neq \bar{G}_{b1}^-(t)$ for all $t>0$. Although such a difference in the individual + and − parallel mean RTs is not sufficient to prove that the models are PST-distinguishable, we will now show that it is sufficient to prove that they *almost* always are. Readers interested in the main track of the development on a first reading may skip to the next proposition.

To keep matters from becoming too complex, suppose that the distribution functions, and therefore the survivor functions, have at most a single point where they are equal, that is, $\bar{G}_{a1}^+(t^*)=\bar{G}_{a1}^-(t^*)$ but $\bar{G}_{a1}^+(t) \neq \bar{G}_{a1}^-(t)$ for $t \neq t^*$, and $\bar{G}_{b1}^+(t^{**})=\bar{G}_{b1}^-(t^{**})$ but $\bar{G}_{b1}^+(t) \neq \bar{G}_{b1}^-(t)$ for $t \neq t^{**}$. Now recall the initial forced equality (from CI 1 + CI 2 = CII 4 + CIII 1),

$$\int_0^\infty [\bar{G}_{a1}^+(t)-\bar{G}_{a1}^-(t)][\bar{G}_{b1}^-(t)-\bar{G}_{b1}^+(t)]\,dt=0$$

Let $\Delta A(t)=[\bar{G}_{a1}^+(t)-\bar{G}_{a1}^-(t)]$ and $\Delta B(t)=[\bar{G}_{b1}^-(t)-\bar{G}_{b1}^+(t)]\,dt$. Now the left-hand side can be decomposed into

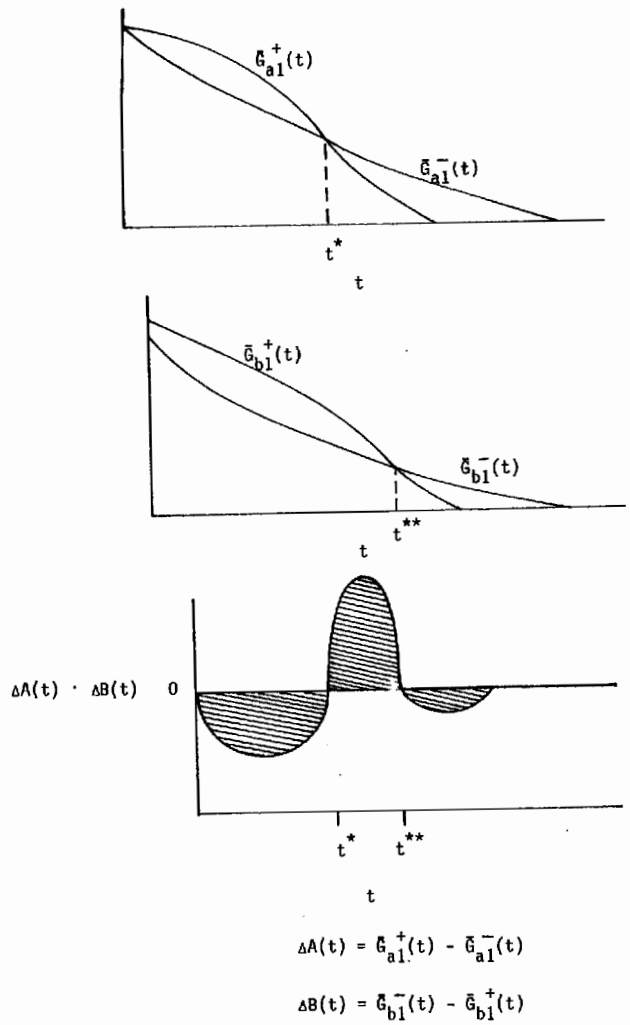


Fig. 13.1. Conditions required for mean observable equivalence in independent parallel models.

$$\int_0^{t^*} [\Delta A(t)\cdot\Delta B(t)] + \int_{t^*}^{t^{**}} [\Delta A(t)\cdot\Delta B(t)] + \int_{t^{**}}^{\infty} [\Delta A(t)\cdot\Delta B(t)]$$

and the middle term differs in sign from the other two, so it must exactly equal the sum of the other two integrals in order to maintain the equality with 0. The situation is illustrated in Fig. 13.1 for a particular hypothetical set of curves; the figure is meant for heuristic purposes, and the areas in the bottom of the figure are inexact. The above-zero area must exactly equal the sum of

the below-zero areas. It would be bizarre if such an outcome happened in nature very often. This is the reason we said that in realistic situations simply having the individual matching vs. mismatching means differ will usually be enough to produce distinct parallel and serial PST predictions, even in the presence of serial position effects.

The next proposition demonstrates the rather obvious statement that the serial model and parallel gestalt model are PST observable mean diverse.

Proposition 13.3: If, in CII 1 and CIII 1, the expected processing time is

$$E(\mathbf{T}_{ab}^{++}) = \int_0^{\infty} \bar{G}_{ab}^{++}(t)\, dt$$

then a parallel model that is gestalt in these (double match) conditions is formed that is observable mean diverse from the serial model. If, exactly as in Proposition 13.2, the first-stage times differ according to match vs. mismatch, the models also differ in structure in the same way that they did in that proposition.

Proof: It is obvious that neither the ordinary parallel nor the serial model can make the strong prediction that CII 1 = CIII 1 as does this gestalt parallel model. Furthermore, with the remainder of the predictions being undamaged (i.e., CII 1 and CIII 1 are not employed in Propositions 13.1 and 13.2), the present parallel and the serial model can be shown to differ by the same method of proof employed in Proposition 13.2. □

Thus, we might expect the present gestalt parallel model to be even more experimentally differentiable from serial models than the regular parallel models since the earlier discriminating aspects of Proposition 13.2 are supplemented by the gestalt differences on conditions CII 1 and CIII 1. This formulation of the gestalt model represents, in a sense, the strongest type of parallel processing, for there is a perfect simultaneous correlation between the processing times of the two elements.

Models based on exponential intercompletion times and examples

The special case where the parallel models are composed of exponential intercompletion times was put forth in an earlier article that focused on the general discriminability of such models from serial models (Townsend 1976a). Even there, the serial models covered in PST were of equal generality to the serial models considered here.

Models based on exponential intercompletion times should not be sneered at. For instance, in its full generality the parallel exponential model possesses 8 free parameters (the $v_{ij}^{\bullet}$, where again $\bullet$ denotes + or −; $i = a, b$; $j = 1, 2$, in Table 13.3) relative to the 10 degrees of freedom in the paradigm, and can exhibit sundry behavioral characteristics of latency – for instance, positive or

Table 13.3. *Serial and parallel observable mean RT predictions for the parallel-serial testing (PST) paradigm*

	Serial models	Parallel models
CI	1. $\frac{p}{u_{a1}^{+}}+\frac{1-p}{u_{b1}^{-}}+\bar{t}_s$	$\frac{1}{v_{a1}^{+}+v_{b1}^{-}}+\bar{t}_p$
	2. $\frac{p}{u_{a1}^{-}}+\frac{1-p}{u_{b1}^{+}}+\bar{t}_s$	$\frac{1}{v_{a1}^{-}+v_{b1}^{+}}+\bar{t}_p$
CII	1. $p\left(\frac{1}{u_{a1}^{+}}+\frac{1}{u_{b2}^{+}}\right)+(1-p)\left(\frac{1}{u_{b1}^{+}}+\frac{1}{u_{a2}^{+}}\right)+\bar{t}_s$	$\frac{1}{v_{a1}^{+}+v_{b1}^{+}}+\frac{v_{a1}^{+}}{v_{b2}^{+}(v_{a1}^{+}+v_{b1}^{+})}+\frac{v_{b1}^{+}}{(v_{a1}^{+}+v_{b1}^{+})v_{a2}^{+}}+\bar{t}_p$
		Gestalt: $\frac{1}{v_g}+\bar{t}_p$
	2. $p\left(\frac{1}{u_{a1}^{+}}+\frac{1}{u_{b2}^{-}}\right)+\frac{1-p}{u_{b1}^{-}}+\bar{t}_s$	$\frac{1}{v_{a1}^{+}+v_{b1}^{-}}+\frac{v_{a1}^{+}}{v_{b2}^{-}(v_{a1}^{+}+v_{b1}^{-})}+\bar{t}_p$
	3. $\frac{p}{u_{a1}^{-}}+(1-p)\left(\frac{1}{u_{b1}^{+}}+\frac{1}{u_{a2}^{-}}\right)+\bar{t}_s$	$\frac{1}{v_{a1}^{-}+v_{b1}^{+}}+\frac{v_{b1}^{+}}{v_{a2}^{-}(v_{a1}^{-}+v_{b1}^{+})}+\bar{t}_p$
	4. $\frac{p}{u_{a1}^{-}}+\frac{1-p}{u_{b1}^{-}}+\bar{t}_s$	$\frac{1}{v_{a1}^{-}+v_{b1}^{-}}+\bar{t}_p$
CIII	1. $\frac{p}{u_{a1}^{+}}+\frac{1-p}{u_{b1}^{+}}+\bar{t}_s$	$\frac{1}{v_{a1}^{+}+v_{b1}^{+}}+\bar{t}_p$
		Gestalt: $\frac{1}{v_g}+\bar{t}_p$
	2. $\frac{p}{u_{a1}^{+}}+(1-p)\left(\frac{1}{u_{b1}^{-}}+\frac{1}{u_{a2}^{+}}\right)+\bar{t}_s$	$\frac{1}{v_{a1}^{+}+v_{b1}^{-}}+\frac{v_{b1}^{-}}{v_{a2}^{+}(v_{a1}^{+}+v_{b1}^{-})}+\bar{t}_p$
	3. $p\left(\frac{1}{u_{a1}^{-}}+\frac{1}{u_{b2}^{+}}\right)+\frac{1-p}{u_{b1}^{+}}+\bar{t}_s$	$\frac{1}{v_{a1}^{-}+v_{b1}^{+}}+\frac{v_{a1}^{-}}{v_{b2}^{+}(v_{a1}^{-}+v_{b1}^{+})}+\bar{t}_p$
	4. $p\left(\frac{1}{u_{a1}^{-}}+\frac{1}{u_{b2}^{-}}\right)+(1-p)\left(\frac{1}{u_{b1}^{-}}+\frac{1}{u_{a2}^{-}}\right)+\bar{t}_s$	$\frac{1}{v_{a1}^{-}+v_{b1}^{-}}+\frac{v_{a1}^{-}}{v_{b2}^{-}(v_{a1}^{-}+v_{b1}^{-})}+\frac{v_{b1}^{-}}{v_{a2}^{-}(v_{a1}^{-}+v_{b1}^{-})}+\bar{t}_p$

negative dependencies among the processed elements. The parallel exponential model provides a quite decent explanation of the data of Townsend and Snodgrass (1974), as we shall see below. An exponential serial model formed a reasonable first approximation to a complex pattern of matching results (Snodgrass and Townsend 1980). Other considerations of model testing will be deferred until the final section.

Table 13.3 exhibits the predictions in the case of exponential intercompletion times. The derivations may be carried out following the rules of Chapters 3 and 4. The serial model predictions, while equivalent to those in Table 13.2, are written in terms of exponential rate parameters for easier comparison with the parallel expressions. It should not be inferred from this tactic that the present serial model is less general than the model of Table 13.2 for predictions of *mean* RT. The symbols for the exponential rates are the same as those used throughout the book, v_{ij} and u_{ij}, where $i=a$ or b, $j=1$ or 2, and $\cdot$ stands for + or −. The relative magnitudes of the rate parameters order the distributions on the intercompletion times so the assumptions about the means and distribution orderings on the parallel models in Propositions 13.1 and 13.2 hold automatically, consequently proving them in the present case. The gestalt exponential parallel model is formulated similarly, with just a single gestalt processing rate applying in both conditions CII 1 and CIII 1.

Observe that the present serial and parallel exponentially based models introduce the additional assumption of across-stage independence (i.e., independence of intercompletion times). Krueger (1978) reported that RT is lower on a particular trial if the preceding trial was a + match rather than a − match. Potentially, this might also occur on a within-trial basis, precluding across-stage independence. In models based on mean RTs, this assumption is unimportant with respect to prediction ability, but might be more critical in attempting to fit RT distribution characteristics.

The above propositions tell us that the parallel and serial models covered can never give identical predictions in PST. However, it is hard to know just how close one model might be able to get to the predictions of the other, in real-life situations, due to the generality and complexity of the paradigm and the attendant models. What we have done to alleviate this problem is to set the parameters of one model to certain values and then to submit the consequent generated PST mean reaction times to a chi-square minimization routine (Chandler 1959), with all the parameters of the alternative model being allowed to vary. There are sometimes difficulties with local minima with this strategy because of the generality of the model being "fit" to the other's predictions. However, reruns with different starting values and careful analysis permit some conclusions to be drawn. Note that we need to test the alternative model in its full generality, for then if it cannot produce a good fit, we have "falsified" a larger set of theoretical alternatives than if we took a narrower case. For convenience, we will call the values generated by the original model the *data* and the values of the best-fitting versions of the alternative model the *predictions*.

The main results are as follows. When degenerate cases were tested, as when no + and − rate differences occur and $u_{a1}=u_{b1}$ (earlier chapters show the parallel and serial models are equivalent everywhere in this instance), the alternate model always predicted the data perfectly. However, the alternate model parameters were typically not those associated with the model mathematically equivalent to the original model, that is, with that model having exactly the same distribution as the data model. This occurred because we are here using only the mean reaction times rather than the entire distribution (see also Theorem 5B in Townsend 1976a: 43).

Secondly, when models meeting the requirements of the propositions above were employed, the alternate model was never able to perfectly predict the data generated by the original model. When the parameters of the alternate model were unconstrained, certain of the rates typically tended toward $+\infty$ or 0, thus devolving the alternate model into a model of the same variety as the original (e.g., a parallel model approaching a serial model) or sometimes a hybrid model.

Letting the residual reaction time parameter increase from 0 ($\bar{t}_s$ in the serial, $\bar{t}_p$ in the parallel model) in the *original* (data) model always permitted the alternate model to get better, with the degree of fit being monotonically related to the size of the residual time parameter. On the other hand, letting the residual time parameter of the *alternate* (predicting) model increase never resulted in better fits. Thus, it appears that the larger the processing component of interest is, relative to the residual component, the better for parallel–serial testability, and the smaller it is, the worse for testing purposes. It also seems that letting $\bar{t}_p$ or $\bar{t}_s$ increase when fitting a wrong model to data has, if anything, a deleterious effect on the fit. Put another way, the mimicking ability of an incorrect model is adversely affected by increasing its residual time parameter. This behavior during a fit by itself might be of aid in rejecting an inappropriate model.

Interestingly, over the range of the models studied in this way so far, the serial model had more difficulty fitting parallel data than vice versa. This seems a little counterintuitive at first, because the serial model possesses more parameters than the parallel model. In the models specifically examined below, p was set to $\frac{1}{2}$ in the serial data models, but this result held true for a wide range of p. It should be kept in mind, however, that the models used to generate data were well-known intuitive ones rather than the most general but sometimes less interesting versions. On the other hand, the particular version of the generating model turned out not to be critical; rather, the overall difference of the + to the − processing rates was the preeminent factor. For instance, the general serial model did not predict data generated by a reallocation parallel model substantially better than it predicted data generated by an independent parallel model (see also Chapters 4 and 6). Rather, the key factor was the disparity of the + and − processing rates in the parallel model. For example, having two large + rates and two small − rates always led to worse fits than one large and one medium + rate and two small − rates.

Table 13.4(a). *Generating model: serial with no serial position effects and* $u^+=u^-$

		Serial data	Parallel predictions	Generating serial parameters	Estimated parallel parameters
CI	1.	2,000	2,000	$\bar{t}_s=1{,}000$	$\bar{t}_p=.02$
	2.	2,000	2,000	$p=.5$	
CII	1.	3,000	3,000	$\frac{1}{u^+}=\frac{1}{u^-}=1{,}000$	$\frac{1}{v_{a1}^+}=6{,}667$
	2.	2,500	2,500		$\frac{1}{v_{a1}^-}=6{,}667$
	3.	2,500	2,500		$\frac{1}{v_{b1}^+}=2{,}866$
	4.	2,000	2,000		$\frac{1}{v_{b1}^-}=2{,}866$
CIII	1.	2,000	2,000		$\frac{1}{v_{a2}^+}=716$
	2.	2,500	2,500		$\frac{1}{v_{a2}^-}=716$
	3.	2,500	2,500		$\frac{1}{v_{b2}^+}=1{,}656$
	4.	3,000	3,000		$\frac{1}{v_{b2}^-}=1{,}656$

Fit: $\chi^2=0$

We now turn to a few examples to illustrate some of these points. Before giving these, however, we have to warn that the absolute magnitudes of the chi-square values are not in themselves meaningful because simply multiplying the theoretical and "observed" values by a constant K multiplies the final chi-square value by K:

$$\chi_k^2=\sum\frac{(KE-KO)^2}{KE}=K\chi_1^2$$

Nonetheless, this technique can show us how close one model's predictions can come to another's, relative to the extremity of the parameter values of the model being fit.

For the sake of simplicity, the models used in the example will not predict serial position effects. The presence or absence of serial position effects appears unrelated to the ability of the opposing, alternate model to approximate the generated data. All of the following cases constrained the "fitting"

Table 13.4(b). *Generating model: serial with no serial position effects and* $u^+>u^-$

		Serial data	Parallel predictions	Generating serial parameters	Estimated parallel parameters
CI	1.	1,250	1,238	$\bar{t}_s=500$	$\bar{t}_p=0$
	2.	1,250	1,238	$p=.5$	
CII	1.	1,500	1,548	$\frac{1}{u^+}=500$	$\frac{1}{v_{a1}^+}=2{,}066$
	2.	1,750	1,780	$\frac{1}{u^-}=1{,}000$	$\frac{1}{v_{a1}^-}=3{,}093$
	3.	1,750	1,780		$\frac{1}{v_{b1}^+}=2{,}066$
	4.	1,500	1,546		$\frac{1}{v_{b1}^-}=3{,}093$
CIII	1.	1,000	1,032		$\frac{1}{v_{a2}^+}=517$
	2.	1,500	1,444		$\frac{1}{v_{a2}^-}=901$
	3.	1,500	1,444		$\frac{1}{v_{b2}^+}=517$
	4.	2,500	2,450		$\frac{1}{v_{b2}^-}=901$

Fit: $\chi^2=1.04$

rate parameters to be greater than .00001 and less than .1, corresponding to times of 10 to 100,000 time units.

The first example, Table 13.4(a), illustrates a case where no + and − processing rate differences occur so that it is expected that the alternate model can predict the data perfectly. Here, the generating model is a standard serial model with $p=\frac{1}{2}$ and $\bar{t}_s=1{,}000$ (the reader may think of the numbers as given in milliseconds), with the rate of processing $u=\frac{1}{1000}$. It can be seen that the parallel model indeed fits perfectly, albeit not with estimated parameters that yield a mathematically equivalent reaction time *distribution*. Note also that the estimated $\bar{t}_p$ was virtually 0, indicating that nonzero values of $\bar{t}_p$ are deleterious to the fitting ability of the parallel model.

Table 13.4(b) shows what happens when the + and − rates differ somewhat and $\bar{t}_s$ is smaller than in Table 13.4(a). It is clear that the parallel model can no longer fit perfectly, although with the usual experimental variances it is not obvious that these two models could be experimentally discriminated

Table 13.4(c). *Generating model: serial with no serial position effects and* $u^+ \gg u^-$

		Serial data	Parallel predictions	Generating serial parameters	Estimated parallel parameters
CI	1.	650	595	$\bar{t}_s = 100$	$\bar{t}_p = 0$
	2.	650	595	$p = .5$	
CII	1.	300	403	$\frac{1}{u^+} = 100$	$\frac{1}{v_{a1}^+} = 787$
	2.	1,150	1,187	$\frac{1}{u^-} = 1{,}000$	$\frac{1}{v_{a1}^-} = 2{,}453$
	3.	1,150	1,187		$\frac{1}{v_{b1}^+} = 787$
	4.	1,100	1,226		$\frac{1}{v_{b1}^-} = 2{,}453$
CIII	1.	200	393		$\frac{1}{v_{a2}^+} = 100$
	2.	700	598		$\frac{1}{v_{a2}^-} = 787$
	3.	700	598		$\frac{1}{v_{b2}^+} = 100$
	4.	2,100	2,007		$\frac{1}{v_{b2}^-} = 787$

Fit: $\chi^2 = 186$

except through attendant considerations, such as number of parameters (here 4 vs. 8 or 9 depending on whether $\bar{t}_p$ was set equal to 0 at the outset) and simplicity or elegance of the model. (Another such consideration will be mentioned later.) On the other hand, with regard to the number of parameters, the parallel number can be effectively reduced to 4 by setting $v_{a1}^+ = v_{b1}^+$ and so on.

We proceed to Table 13.4(c) to offer a case where u^+ is substantially larger than u^- and $\bar{t}_s$ is also smaller than before. The parallel predictions are rather poor, although they more or less follow the trend of the serial generated data. If we had set $\bar{t}_p = 0$ before carrying out the fit and then set $v_{a1}^+ = v_{b1}^+$ and so forth, we would have only three parallel parameters, less than the number of serial parameters, 4. In the present instance, we would probably still choose the serial model, however, because of the much better fit it would give, but this example sounds a warning of perils surrounding quick conclusions following model fits.

The next example takes up a parallel generating model [Table 13.5(a)] with

Table 13.5(a). *Generating model: parallel with no serial position effects and* $v^+ > v^-$

		Parallel data	Serial predictions	Generating parallel parameters	Estimated serial parameters
CI	1.	833	851	$\bar{t}_p = 500$	$\bar{t}_s = .10$
	2.	833	851	$\frac{1}{v^+} = 500$	$p = .43$
CII	1.	1,250	1,187	$\frac{1}{v^-} = 1{,}000$	$\frac{1}{u_{a1}^+} = 572$
	2.	1,500	1,435		$\frac{1}{u_{a1}^-} = 760$
	3.	1,500	1,435		$\frac{1}{u_{b1}^+} = 920$
	4.	1,000	932		$\frac{1}{u_{b1}^-} = 1{,}062$
CIII	1.	750	770		$\frac{1}{u_{a2}^+} = 366$
	2.	1,000	1,060		$\frac{1}{u_{a2}^-} = 1{,}025$
	3.	1,000	1,060		$\frac{1}{u_{b2}^+} = 484$
	4.	2,000	2,100		$\frac{1}{u_{b2}^-} = 1{,}355$

Fit: $\chi^2 = 26.89$

a moderate difference of + and − processing rates and $\bar{t}_p = 500$. The serial predictions are not far off but are not too good either. Further, the number of serial parameters cannot be reduced here (except to let $\bar{t}_s \doteq 0$) so that both fit and parsimony would argue for the parallel model. However, it might be that a constrained serial model with no *a* vs. *b* rate differences would perform nearly as well, because $v_a^\bullet = v_b^\bullet$ ($\bullet = +, -$) in the parallel model. This has not been tried.

The final example employs the same kind of parallel model but with more disparate processing rates and a smaller residual parameter, $\bar{t}_p = 100$. Table 13.5(b) exhibits the inability of the serial model to closely approximate the parallel data, and it appears that most of the serial parameters are needed to do as well as it does.

A shortcoming of the above results is that they ignore variances and the statistical aspects of hypothesis testing. This could be alleviated by working out the variances of the processing times and positing some distribution on $\bar{t}_s$

Table 13.5(b). *Generating model: parallel with no serial position effects and* $v^+ \gg v^-$

		Parallel data	Serial predictions	Generating parallel parameters	Estimated serial parameters
CI	1.	191	274	$\bar{t}_p = 100$	$\bar{t}_s = .10$
	2.	191	274	$\frac{1}{v^+} = 100$	$p = .53$
CII	1.	250	199	$\frac{1}{v^-} = 1{,}000$	$\frac{1}{u_{a1}^+} = 10$
	2.	1,100	986		$\frac{1}{u_{a1}^-} = 278$
	3.	1,100	986		$\frac{1}{u_{b1}^+} = 269$
	4.	600	414		$\frac{1}{u_{b1}^-} = 566$
CIII	1.	150	133		$\frac{1}{u_{a2}^+} = 70$
	2.	200	307		$\frac{1}{u_{a2}^-} = 1{,}501$
	3.	200	307		$\frac{1}{u_{b2}^+} = 63$
	4.	1,600	1,840		$\frac{1}{u_{b2}^-} = 1{,}356$

Fit: $\chi^2 = 280$

and $\bar{t}_p$. However, whether this would be of immediate value is open to question, since we know so very little about the residual time distribution a priori. As noted above, we can, at least in the present approach, acquire some idea of how closely one model can approximate the mean reaction time data predicted by another in an ideal sense. Put another way, the above examples represent what might be expected if an extremely large number of trials were run so that the mean reaction times approached their expected values.

There are several conclusions we may draw from the above examples and related concerns. The first is that PST works, at least ideally, and secondly the indistinguishability of ordinary parallel and serial models in the absence of special discriminating circumstances (e.g., when $u^+ = u^-$) is again pointed out. On the other hand, in some cases, such as if the residual reaction time is large relative to the processing rates and if there is only a small + and − rate difference, an alternate model may give predictions that are reasonably close to the true model.

How can the experimenter optimize the chances for discriminability? She

or he should design the study so as to maximize the + and − processing time difference, particularly relative to the presumed residual time magnitude. Parsimony or elegance can also help to support one of the models over another, as when one has significantly more parameters. Another important observation, which is illustrated in the above examples and not emphasized previously, is that when the best-fitting model of one variety (e.g., parallel) has a ridiculously small residual time parameter estimate while the other has a reasonable residual parameter, support is clearly gained for the latter.

It may be possible to strengthen PST by comparing distributional predictions with the actual observed distributions. However, this may lose as much in generality as it gains in sensitivity. Perhaps a preferable way to enhance testability (but a way not mutually exclusive with the first) is to utilize the *principle of correspondent change,* which says the correct model should evidence the proper type of parameter alterations or invariances across planned experimental manipulations, whereas the incorrect model should not (e.g., see chapter 15).

For example, one use of the principle would be to experimentally vary the rate, say, of the positive matches. In a visual realization of PST this could be accomplished by manipulating the visual intensity of the comparison element matching the target. If the correct model were serial, then only u^+ would show much alteration across these conditions, whereas the parallel model would evince changes in both v^- as well as v^+. Calculations such as those illustrated in Tables 13.4 and 13.5 show that this can be expected to happen, although as could also be expected, when u^+ alone changes, the estimated v^+ parameter reveals more variation than does v^-.

Another experimental change that might be expected to alter the + or − rate only in the correct model would be to make matches more or less frequent than mismatches, as suggested by Krueger (1978) in a reanalysis of a study by Coltheart and Curthoys (1968). On the other hand, varying the frequency of a match occurring in, say, position a, would be somewhat less revealing. Basically, if processing were serial, only p should change. However, v_{aj}^* $(j = 1, 2)$ might be expected to vary relative to v_{bj}^* $(j = 1, 2)$ whether or not processing were parallel. Thus, in this instance, a serial finding might be more impressive.

In summary, the present investigation has given us some indication concerning the ability of models that have been shown to be analytically separable via a special experimental paradigm to approximate one another. It has also helped to suggest ways of optimizing the discriminability of parallel and serial models vis-à-vis PST.

An application

To our knowledge, there has been only one application of the PST paradigm to date, and that is one carried out in a collaborative effort of the first author with Professor Joan G. Snodgrass at New York University (Townsend &

Snodgrass 1974). The study employed a visual target consisting of two characters presented side by side, while the comparison stimulus was also two characters each one of which appeared just below one of the target stimuli. The observer's task was to compare the vertically displaced characters on the right with one another and to similarly compare the vertically displaced characters on the left with one another; no diagonal comparisons were necessary.

In condition CI 1, an R2 response was required because the two characters on the left matched (were physically identical) whereas the two characters on the right were different. The opposite situation was the case in CI 2. The logical conjunction demanded in CII requires that the pairs of stimuli on the right as well as on the left must match in order for the R1 response to be required; otherwise the observer makes the R2 response. Condition CIII, on the other hand, permits an R1 response when either the pair on the right or the pair on the left match, and the R2 response is to be made only when neither pair is a + match. These and the other conditions of the study are illustrated in Table 13.6.

We now describe the experiment in more detail and then return to the results.

PST experiment: method

Observers. Four observers served in the experiment, of whom two were females and two male. One female and one male were graduate students, while the others were undergraduates. All were right-handed and all were paid for their participation.

Stimuli. The stimuli were all possible ordered pairs of the four upper-case letters *Q*, *R*, *T*, and *Z* (with no repeated letters). The stimuli were photographed from letters mounted on white cards (Instantype L-1510) and made into 35-mm black-and-white slides. As viewed in the tachistoscope, the letter pairs subtended 3.5 deg of visual angle vertically and between 5.5 deg and 9 deg horizontally.

Apparatus. The stimulus pairs were presented sequentially via a Scientific Prototype automatic three-channel tachistoscope (model GB) equipped with a binocular zoom lens, Kalimar K 7012. The stimulus durations, interstimulus interval, and intertrial interval were controlled automatically by three time interval generators. The first stimulus pair was presented for 300 msec, followed by a 400-msec blank lighted field, followed by the second stimulus pair that was exposed for 300 msec. There was a 4,000-msec intertrial interval timed from the offset of the second stimulus (rather than from the subject's response).

The onset of the second stimulus started an electronic counter (Montsanto Counter-Time #101b), and a press of one of two response keys by the observer stopped the timer and displayed the RT to the nearest msec. A light on the experimenter's console displayed which response key the observer had pushed.

Table 13.6. *Description of each condition with total number of trials per trial type per observer and mean reaction time depicted*

Condition		Examples of stimuli	Correct response	Number of trials	Average (over observers) RT
CI	1.	Target *QT* Comparison *ZT*	R2	720	381
	2.	Target *QT* Comparison *QZ*	R1	720	356
CII	1.	Target *QT* Comparison *QT*	R1	2,160	347
	2.	Target *QT* Comparison *QZ*	R2	720	400
	3.	Target *QT* Comparison *ZT*	R2	720	394
	4.	Target *QT* Comparison *ZR*	R2	720	367
CIII	1.	Target *QT* Comparison *QT*	R1	720	352
	2.	Target *QT* Comparison *QZ*	R1	720	437
	3.	Target *QT* Comparison *ZT*	R1	720	440
	4.	Target *QT* Comparison *ZR*	R2	2,160	464

The observer was seated at a table and viewed the stimuli through the binocular zoom lens that was provided with rubber eye cups. Each eyepiece was focused independently by each observer before each session. The experimenter was seated behind the tachistoscope, in the same room with the observers, and started and stopped the stimulus presentations, recorded RTs, and informed the observer when he or she had made an error. Automatic changers in both fields advanced slide trays (Sawyers Rototray) containing the sequence of stimuli for each session.

Design and procedure. Each observer participated in 3 practice and 30 experimental sessions. Each of the three conditions (CI, CII, and CIII) were run in separate practice and experimental sessions rotated in a balanced design so that practice efforts would be equally distributed across the three main conditions. Because we wanted to equalize the number of possible trial types within a condition, more trials were run with the more complex condi-

tions CII and CIII than with CI. The number of trials for CI was 48, and for CII and CIII 144 for each session. Table 13.6 summarizes the number of trials per type per observer within all 30 sessions.

For each condition in each session, the number of R1 and R2 trials were equal, all possible pairs occurred equally often, and the spatial locations of matching stimuli were counterbalanced across right and left positions. In addition, matching (i.e., identical) stimuli never occurred on opposite sides of the stimulus, as shown in Table 13.6.

A typical session lasted approximately one hour. Prior to running the experimental trials for each condition, 10 practice trials, selected randomly from the experimental sequence, were run in order to familiarize the observer with the condition. Observers were paid \$1.50 for participating in each session and in addition won money for fast correct responses and were penalized for errors. Each observer responded R1 ("right" or "same") with the right hand, and R2 ("left" or "different") with the left.

Observers were instructed to attempt to use a visual matching strategy in making their responses and to avoid verbalizing the stimuli to themselves if possible. They were also instructed in the differences between parallel and serial processing and were interviewed after each session about which of the two strategies they thought they had used. Observers did not always report using a uniform strategy with respect to verbalization across the four conditions, and in addition they did not report consistently feeling they were using either parallel or serial search; the reported search strategy appeared to depend upon the difficulty of the condition (with difficulty increasing with condition number) and amount of practice.

Results and discussion

Table 13.6, which illustrates the basic realization of the design, also reports the overall mean reaction times for the various conditions, averaged across observers. Cursory inspection reveals that CII 1 = CIII 1, strongly suggesting a gestalt perception of the double matches. On the other hand, remaining conditions lead to the tentative inference (see Chapter 7 for proper caveats) that self-termination occurs because, for example, CIII 4, which required exhaustive processing, is greater than the other CIII conditions, and CII 2 and 3, which sometimes demanded the processing of both comparisons, are greater than CII 4, which only required the processing of a single comparison. A final qualitative observation is that overall, under the postulates of the method, it appears that the mismatches might take longer on the average than matches since the CIII times tend to be longer than those in the other conditions.

The three basic models of Table 13.3 were fit to the data, employing Chandler's (1959) STEPIT routine for minimization of χ^2. Table 13.7 shows the observed and predicted values as well as the χ^2 values associated with the fits. The value of P = probability of a χ^2 value being equal to or greater than

Table 13.7. *Model predictions for individual observers in msec*

		Observer 1				Observer 2			
		Obs.	Ser.	Reg. par.	Ges. par.	Obs.	Ser.	Reg. par.	Ges. par.
CI	1.	383	383	381	381	338	342	338	338
CI	2.	348	357	350	356	312	315	312	316
CII	1.	346	370	372	339	314	345	314	315
CII	2.	429	415	428	437	348	353	348	348
CII	3.	411	397	411	411	338	342	338	338
CII	4.	387	403	405	381	321	349	321	316
CIII	1.	332	336	332	339	316	308	316	315
CIII	2.	397	383	381	397	372	342	372	371
CIII	3.	415	391	390	407	382	352	382	382
CIII	4.	459	476	459	459	393	387	393	393
χ^2(df)			6.1(0)	5.0(2)	.83(2)		10.7(0)	1.29(2)	.134(2)
P			—	.082	.667		—	.529	.93

Table 13.7 *(cont.)*

		Observer 3				Observer 4			
		Obs.	Ser.	Reg. par.	Ges. par.	Obs.	Ser.	Reg. par.	Ges. par.
CI	1.	435	453	459	440	367	360	352	360
CI	2.	368	386	368	389	398	391	398	392
CII	1.	364	451	397	388	362	411	376	357
CII	2.	445	453	459	440	379	398	379	382
CII	3.	441	449	441	441	385	404	399	392
CII	4.	409	453	409	389	351	408	352	355
CIII	1.	410	386	381	388	352	342	352	357
CIII	2.	511	453	475	510	467	422	467	469
CIII	3.	523	451	523	523	438	396	423	438
CIII	4.	524	516	524	524	479	458	479	477
χ^2(df)			43.43(0)	9.36(1)	2.73(1)		26.0(0)	2.19(1)	.577(2)
P			—	.003	.094		—	.135	.753

that obtained is also shown except for the serial model whose 10 estimated parameters left 0 degrees of freedom. Before discussing the parallel models, it can be readily seen that the serial models were convincingly falsified since the χ^2 values were large in all cases despite there being as many free parameters as data points. This result points up the fact that PST carries substantial power for testing processing structure; the more general the falsified model the better. The failure of the serial models is also demonstrated in Table 13.8, which gives the parameter estimates of the serial model, because it can be seen that every observer evidences at least one absurd parameter value. The zeros in Table 13.8 correspond to infinite processing rates and suggest that there must be some parallelity in the data.

Although the prediction that CI 1+CI 2=CII 3+CIII 4 was not too far off the mark in an absolute sense, the parallel models were closer to correctly predicting the separate sums CI 1+CI 2 and CII 3+CIII 4. Another major disparity was in the serial prediction that CII 1 should be much greater than CIII 1, a disparity that helped to force the serial models to place most of the processing duration variability in u_{b1}^{+} and u_{b1}^{-}.

The most general regular parallel model possesses 9 free parameters, and the most general gestalt parallel model had 10 free parameters. Despite the 10 parameters, the gestalt model is not guaranteed a perfect fit. For instance, *none* of the present models could accommodate the result CII 2<CI 1. The general regular parallel model is also certainly falsifiable even with all 9 parameters, but some data snooping suggested some more constrained models that, relative to their degrees of freedom, performed about as well as the original general model. The more constrained models were generated by setting certain of the parameters equal to one another, as indicated in Table 13.8 where the constrained parameters are underlined. These models produced the results of Table 13.7.

The regular parallel models appear to handle the data somewhat better than the serial model, when the additional dfs are taken into account. Note that the regular parallel models tended to handle the result CII 1≅CIII 4 better than the serial models, although a disparity remains. The gestalt models are best for all observers, with the fits being excellent for all except observer 3, whose fit was only moderate. The gestalt parameter estimates are generally reasonable, but the estimated $\bar{t}_p=0$ in the case of observer 3 and $v_{b1}^{+}=0$ for observer 4 are distressing. For some reason, all the play in the model-processing structure is needed for observer 1 to model the data well, forcing $\bar{t}_p$ to 0. A very speculative surmise is that this observer evidences more temporal overlap among his or her internal processes.

Observer 3 may be carrying out a type of hybrid processing in which positive matches in the a (left) position are virtually ignored until the b position match is completed (except, of course, on the gestalt stimuli). A similar explanation may be appropriate in observer 4's behavior.

Three of the four observers seemed to process $+a$ matches at about the same rate as $-a$ matches, when this position was completed second (observers

Table 13.8. *Parameter estimates of models fit in PST application*

Observer	p	$\bar{t}_s$	Serial								Regular parallel							
			$\frac{1}{u_{a1}^+}$	$\frac{1}{u_{a2}^+}$	$\frac{1}{u_{b1}^+}$	$\frac{1}{u_{b2}^+}$	$\frac{1}{u_{a1}^-}$	$\frac{1}{u_{a2}^-}$	$\frac{1}{u_{b1}^-}$	$\frac{1}{u_{b2}^-}$	$\bar{t}_p$	$\frac{1}{v_{a1}^+}$	$\frac{1}{v_{a2}^+}$	$\frac{1}{v_{b1}^+}$	$\frac{1}{v_{b2}^+}$	$\frac{1}{v_{a1}^-}$	$\frac{1}{v_{a2}^-}$	$\frac{1}{v_{b1}^-}$	$\frac{1}{v_{b2}^-}$
1	.51	203	0	0	273	67	40	84	369	63	0	387	0	2,314	48	412	405	24,743	48
2	.67	217	0	0	272	56	10	83	375	16	312	∞	12	0	80	0	182	36	80
3	.34	215	0	0	258	191	0	94	358	0	315	∞	16	66	796	273	91	144	161
4	.48	193	0	119	289	11	99	25	322	78	352	0	367	∞	24	47	127	0	39

Observer	Gestalt parallel											
	$\bar{t}_p$	$\frac{1}{v_g}$	$\frac{1}{v_{a1}^+}$	$\frac{1}{v_{a2}^+}$	$\frac{1}{v_{b1}^+}$	$\frac{1}{v_{b2}^+}$	$\frac{1}{v_{a1}^-}$	$\frac{1}{v_{a2}^-}$	$\frac{1}{v_{b1}^-}$	$\frac{1}{v_{b2}^-}$
1	0	339	398	364	3,313	58	398	515	8,799	58
2	280	35	769	37	63	157	88	37	63	132
3	250	138	∞	71	190	500	520	71	190	310
4	313	44	99	207	∞	46	79	207	89	46

2, 3, 4). Observers 2 and 3 processed position *b* matches during stage 1 at the same rate whether they were matches or mismatches, and observers 1 and 4 did the same for the *b* position when it was completed second. Finally, stage 1 rates were about equal for observer 1 on position *a*. In the other cases, the + rates were all greater than the − rates for the same stage and position, thus partially confirming the surmise that negative matches consume more time than positive matches. It further appears that more parameters could be set equal to one another without significantly harming the fits.

From the above analyses, we can conclude that substantial parallelity was present and that there was a strong predisposition to process the double + target stimulus as a single unit.

Incidentally, it is intriguing that the gestalt model is easily tested against the alternatives. The strong properties associated with gestalt activity may be one reason a number of cognitive or information-processing investigators are becoming interested in gestalt psychology (e.g., see Navon 1977).

Although the present results strongly suggest parallelity, they do not, of course, tell us how large the visual parallel span of apprehension can be before a switch to serial processing is mandated. Nevertheless, a substantial range of data suggests about four to six unrelated elements may be visually processed in parallel (Hamilton 1859; Krueger 1978; see also particularly Chapters 6 and 11). The actual "glimpse" limit pertinent in any particular task undoubtedly depends on the specific processing demands of that task. For instance, matching of the Estes and Taylor (1964) detection paradigm variety in which an observer must determine whether a *B* or *F* (say) is present in the display may require less visual processing capacity on each displayed item than a whole report task, where as many letters as possible must be actually identified (see Chapter 11). It also need not be presumed that processing is parallel *unlimited capacity* up to the parallel limit, although in gestalt or so-called configurational perception, unlimited capacity is a reasonable axiom.

There are some curious aspects of the model fits that indicate some hybrid processing. The present experiment was run with the various conditions on a between-block basis. This might lead to some attentional or strategy changes by some observers in the separate conditions (as in the Shield conditions of the whole report experiment in Chapter 11). In future research, it should be possible with well-trained observers to employ three cue lights to indicate from trial to trial which condition is being tested. This would increase the probability that the observers' strategies remain relatively constant across conditions.

PST and distributional diversity and testability

When the most general of parallel distributions are introduced (for example, they need not obey within-stage independence), then the serial and parallel models are not necessarily observable mean diverse – the PST mean reaction

times could be equal. On the other hand, as will be seen below, a case can be made that this would be unlikely in practice. A distinct question, but one that will help to explicate the first, is whether the actual probability distribution functions on reaction times can be made to be equivalent when it is assumed that + and − distributional differences are present.

The distribution-free prediction by serial models that CI 1 + CI 2 = CII 4 + CIII 1 clearly continues to hold at the distributional level, if it is assumed that the density on t_s, $h_s(t_s)$, is independent of the comparison time distribution. To see this, let $f_i'(t) = f_i(t) * h_s(t)$, $i = a, b$, that is, the convolution of the residual and comparison time distributions. In terms of the serial survivor functions we then have

$$\text{CI 1} + \text{CI 2} = \text{CII 4} + \text{CIII 1}$$
$$= p[\bar{F}'_{a+}(t) + \bar{F}'_{a-}(t)] + (1-p)[\bar{F}'_{b+}(t) + \bar{F}'_{b-}(t)] \quad \text{for all } t \geqslant 0$$

The stage subscript has been dropped. Our remarks will emphasize this relation, which was important in the earlier distinguishability theorems at the level of the means, so that we can ignore stage 2 distributions. Similar comments can be made about other aspects of the PST paradigm.

We now require a notation for the parallel first-stage joint distributions. Consider for exemplary purposes the instance where the a position element matches the target but the b position element does not. Then the survivor function for the minimum completion time will be expressed as

$$P(\mathbf{T}_{a1}^{+} \geqslant t \cap \mathbf{T}_{b1}^{-} \geqslant t) = \bar{G}_{a+b-}(t)$$

Let us convolve this with the density on the residual time t_p to get $g'_{a+b-}(t) = g_{a+b-}(t) * h_p(t)$ and the associated $\bar{G}'_{a+b-}(t)$. This expression is appropriate as the survivor function on the completion time in condition CI 1. The other cases will be denoted analogously.

Now, what interpretations should be given the PST demand that + and − processing distributions should be different? As occurs often in probabilistic arenas, there are several possibilities that come to mind. The various alternatives depend on how many time points t the distribution functions differ and on whether there is an ordering at some or all of those times or simply a difference in either direction. It seems reasonable to suppose an ordering of the distribution functions because that corresponds to a probabilistic version of, say, positive processing times being shorter than negative processing. Suppose, then that we posit an ordering of the form

$$\bar{G}_{a+b+}(t) < \left\{ \begin{matrix} \bar{G}_{a+b-}(t) \\ \sim \\ \bar{G}_{a-b+}(t) \end{matrix} \right\} < \bar{G}_{a-b-}(t)$$

where the $\sim$ in the middle brackets means the order between $\bar{G}_{a+b-}$ and $\bar{G}_{a-b+}$ is open. It is easy to show that convolution preserves order in distribution or survivor functions so the above order continues to hold on the overall observed RT survivor functions. We still have to decide on what number of values of t this ordering should hold. Natural alternatives are (i) for all $t > 0$;

(ii) for all but a finite number of values; (iii) for at least one value of t. The third possibility is certainly too lenient to be very meaningful. Either (i) or (ii) seems reasonable, and for convenience let us look at (i): The ordering should hold for all $t > 0$. The identity issuing from seriality is, in terms of the parallel functions,

$$\left[\begin{matrix} \text{CI 1} + \text{CI 2} = \text{CII 4} + \text{CIII 1} \\ \bar{G}'_{a+b-}(t) + \bar{G}'_{a-b+}(t) = \bar{G}'_{a+b+}(t) + \bar{G}'_{a-b-}(t), \quad t \geqslant 0 \end{matrix} \right]$$

Now, analytically the ordering given above on the $\bar{G}$ terms can hold and still satisfy this equality since both $\bar{G}'_{a-b+}$ and $\bar{G}'_{a+b-}$ lie between $\bar{G}'_{a+b+}$ and $\bar{G}'_{a-b-}$ for all $t > 0$. However, the point can be made that there is no good reason that nature would produce such an anomaly; we would surely expect that there would exist large ranges of time where the left-hand side would not equal the right-hand side. Taking enough trials should test the prediction. There are some not unreasonable situations where more can be said. For instance, the left-hand side is always greater than the right-hand side if increasing the number of negative matches from one to two has a uniformly bigger effect than increasing the number of negative matches from zero to one; that is, $\bar{G}'_{a-b-}(t) - \bar{G}'_{a+b-}(t) > \bar{G}'_{a-b+}(t) - \bar{G}'_{a+b+}(t)$, because it then follows that CI 1 + CI 2 < CII 4 + CIII 1 for all $t > 0$, expressed in terms of the survivor functions. This prediction can then be tested by a one-sided Kolmogorov test by first dividing both sides by 2 and then converting each into a survivor function.

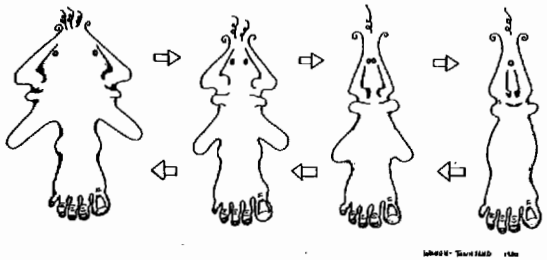
Interestingly, the above condition is met when the intercompletion times are independent, because then $\bar{G}'_{a+}(t) < \bar{G}'_{a-}(t)$, $\bar{G}'_{b+}(t) < \bar{G}'_{b-}(t)$ and

$$\text{CI 1} + \text{CI 2} - [\text{CII 4} + \text{CIII 1}] = [\bar{G}'_{a-}(t) - \bar{G}'_{a+}(t)][\bar{G}'_{b-}(t) - \bar{G}'_{b+}(t)] > 0, \quad t > 0$$

proving the assertion. (This conclusion is related in a general way to the first sections of Chapter 12.)

The upshot of this discussion is that at the more sensitive distributional level we can expect to find aspects of PST that discriminate parallel and serial models even when virtually no assumptions are made about the actual distributions. A very weak proviso should be added, however. We need to employ (and do in the present treatment) realistic constraints on the two types of processes (see also Chapter 15). For example, it is not allowed that the serial distribution on the first element processed depend on whether or not the other element matches or mismatches the target. This seems eminently reasonable, for otherwise the serial mechanism must be getting information about the second element before it is processed.

Finally, it can again be emphasized, at the hazard of redundancy, that the principle of correspondent change (Chapter 15) will strengthen testability whenever properly applied. Thus, the parameter p is not expected to depend on the nature of the trial, for instance, whether both elements match or mismatch the target.



The continuous transformation of one being into another and vice versa. In this case, the creature on the left might symbolize a parallel model and the one on the right a serial model.

14 *Stochastic equivalence and general parallel-serial equivalence relations when system differences are minimal or ignored*

This chapter will be devoted to discussing the formulation of equivalence relations when almost none of the aspects that can differentiate parallel and serial processing are taken into account (or if none of them are operative in a given system or class of systems). Many such differentiating aspects have been touched on in previous chapters. However, one may acquire the flavor of such aspects by perusing the beginning of the following chapter, which takes up a general survey of these.

In Chapter 4, a (stochastic) model was defined essentially to be a probability distribution on an event space of interest to the scientist. A class of models was then a set of probability distributions on the space in question. In most cases investigated in earlier chapters, the set of distributions was prescribed by giving one or more functions of a set of parameters. When the parameters were given specific numerical values, one of the admissible distributions was produced, thus generating a model. Of course, a model may "be" the model class if it is the only member of the set of distributions prescribing the class. (In a number of cases, we also referred to a model class as the "model" for linguistic simplicity.) We shall continue this terminology, but in a more general setting. In order to deal effectively with a general notion of model equivalence, it is necessary to provide a terse description of probability spaces and of implication and equivalence.

The reader who is already versed in elementary measure-theoretic concepts of probability theory should skip ahead in this chapter to the section "Equivalence of parallel and serial models."

A synopsis of implication and equivalence relations in probability spaces and models

In starting out, the reader is cautioned that investigation of the most general stochastic processes, including, for instance, those involving both continuous-time and continuous-state spaces, involve rather subtle notions of topology and analysis. Fortunately, we can dispense with the more esoteric productions here. A little naive set theory and some elementary knowledge of mappings is sufficient for our purposes.

First, we require the set of things of which *events* of interest will be subsets. Let this set be called W. For instance, in a series of two coin tosses, $W=\{(H,H),(H,T),(T,H),(T,T)\}$, where H=heads and T=tails. Often, the elements of W are called *points* or *atoms*.

Next enters the concept of the set of potential *events,* or *field.* A field $\mathfrak{B}$ is a set of subsets of W that is closed under one or more set operations, such as intersection, union, and complementation. Thus, if $A_1 \subseteq W$ and $A_2 \subseteq W$ and $A_1 \in \mathfrak{B}$ and $A_2 \in \mathfrak{B}$, then if $\mathfrak{B}$ is a field, it must be that $\{A_1 \cup A_2\} \in \mathfrak{B}$, $\{A_1 \cap A_2\} \in \mathfrak{B}$, $\bar{A}_1 \in \mathfrak{B}$ ($\bar{A}_1$=complement of A_1 with respect to W) and so on.

When the set of events under discussion is infinite, then the field must be a *sigma field,* defined to be a field that is closed under a countable number of set operations (*countable* means that a one-to-one mapping can be set up with some subset of the integers).

In the coin toss example, the usual field is just $\mathfrak{B}=\{\{(H,H)\},\{(H,T)\},\{(T,H)\},\{(T,T)\},\{(H,H),(H,T)\},\{(H,H),(T,H)\},\{(H,H),(T,T)\},\{(H,T),(T,H)\},\{(H,T),(T,T)\},\{(T,H),(T,T)\},\{(H,H),(H,T),(T,H)\},\{(H,H),(H,T),(T,T)\},\{(H,H),(T,H),(T,T)\},\{(H,T),(T,H),(T,T)\},\{(H,H),(H,T),(T,H),(T,T)\},\varnothing\}$, where $\varnothing$=empty set.

Thus, $\{(H,T),(T,T)\}=\{(H,T)\}\cup\{(T,T)\}$. Note that any set operation on members of $\mathfrak{B}$ again results in a member of $\mathfrak{B}$ and that $\{(H,H),(H,T),(T,H),(T,T)\}=W$, the universal set. This particular field can be seen to be the so-called power set – the set of all subsets of W. Because there are 4 elements of W, there are 16 members of $\mathfrak{B}$. The power set is a frequently employed field when W is finite, but, of course, is not the only one possible.

The other major concept required for the definition of a probability space is, reasonably enough, the assignment of probabilities (P) to the members (events) of $\mathfrak{B}$. We shall develop the idea of probability more below, but certainly we desire that $P(A)\geqslant 0$ when $A\in\mathfrak{B}$ and that

$$\sum_{\text{all } A\in\mathfrak{B}} P(A)=P(\text{Union of all sets contained in } \mathfrak{B})=P(W)=1$$

that is, a probability should be greater than or equal to 0 and the sum of probabilities over all potential events should be 1 (in other words, *something* has to happen). Let P refer to the probability function on $\mathfrak{B}$.

We can now define a probability space M as the ordered set $M=\langle W,\mathfrak{B},P\rangle$. Here M is appropriately the title of the model because a "model," in our

sense, is just a probability space. When the underlying W and $\mathfrak{B}$ are understood, it is acceptable to identify a model with P, the actual probability function. Thus, throughout the book, W=positive real "time" axis $\mathbf{R}^+$, and $\mathfrak{B}$ was implicitly and informally taken as the set of "events" on the real line. These concepts will be further clarified below. First, we duly acknowledge the concept of a minimal sigma field, which is essentially the smallest class (set) of sets giving the desired closure with regard to countable set operations. This notion is required for technical reasons that we cannot go into here (see Loève 1955). Further, the sets belonging to the minimal sigma field are said to be measurable because these are the entities to which measures (in our case, a *probability* measure) are assigned.

When W is made up of one or more dimensions each of which consists of a real line, and the sets going into our class of sets are intervals (e.g., an interval in a one-dimensional space open on the right and closed on the left is $[r_1, r_2)$ with $r_1 < r_2$ both real numbers), then the minimal sigma field of interest is conventionally referred to as a *Borel field.* Thus, in the present investigation $W=\mathbf{R}^+$, a real n-dimensional space defined for all positive (or 0) values on the n dimensions and $\mathfrak{B}$=Borel field based on the sets of points contained in W. Any set contained in the Borel field $\mathfrak{B}$ is referred to as a *Borel set,* and thereby is an interval, appropriately defined in $\mathbf{R}_n^+$, or is produced by countable unions, intersections, and/or complementations of intervals. The sets belonging to $\mathfrak{B}$ are called *measurable* or simply Borel sets.

The idea behind all this, which can go into many pages in a reasonably complete treatment, is to produce a probability space that is internally consistent (although ultimately, all probability representations seem to lead to paradoxes; see, e.g., Fine 1973) and conserves probability measure appropriately [e.g., $P(W)=1$, whatever combination of sets in $\mathfrak{B}$ gives the "universe" W]. The sets in $\mathfrak{B}$ are, of course, associated with the events whose probability we are "measuring," in particular, in the present case, with intervals of time in which an element might complete processing (and whose probability of finishing in that interval can be assessed). The doublet consisting of both a set of positive real dimensions, formally the Cartesian product of $\mathbf{R}^+$, and the Borel field coming out of those dimensions is called a *Borel space,* often written $\langle \mathbf{R}_n^+, \mathfrak{B}^n \rangle$.

Next we require the notion of a Borel function J. Naturally, the notion is useful for (and confined to) relations between two Borel spaces. As with our usual idea of functions, the present ones are single-valued, so if we are mapping from a space $S_1=(W_1, \mathfrak{B}_1)$ to $S_2=(W_2, \mathfrak{B}_2)$, each point in the first can have only one counterpart in the second, although any point in the second can be the image (i.e., function) of more than one point in the first. Corresponding to the image is the inverse image, which is just the reverse mapping from S_2 back to S_1. So, by the foregoing, any point in S_2 can map to a subset of points back in S_1, but all of these latter subsets (for separate points in S_2) will be disjoint. One of the consequences of this is that the inverse mapping

preserves all set operations and class inclusion relations found in S_2. This conclusion will be illustrated below.

A Borel function may now be defined as a function having the property that the inverse image of any Borel set in S_2 is a Borel set in S_1. When applied to spaces more general than Borel spaces, they are simply called *measurable functions.*

A special case of a measurable function of considerable import is a *random variable,* which is obviously rather misnamed. For it is really a mapping from W into the set of real numbers $\mathbf{R}$ such that the inverse images of Borel sets of $\mathfrak{B}_\mathbf{R}$, the Borel field of $\mathbf{R}$, are measurable in $\mathfrak{B}_W$ (i.e., are members of $\mathfrak{B}_W$), the minimal sigma field associated with W. However, the important point is that we are associating numbers with events in $\mathfrak{B}_W$. For instance, suppose a dollar is gained or lost depending on the outcome of a coin toss. Then we would define a random variable $\mathbf{D}$ as a function on W with the property that $\mathbf{D}(H)=1$, $\mathbf{D}(T)=-1$. We could thus speak of the expectation of $\mathbf{D}$ as the average earnings on a single coin toss. If p is the probability that a possibly biased coin will turn up heads, then

$$\begin{aligned} E(\mathbf{D}) &= P(\mathbf{D}=1)\cdot 1 + P(\mathbf{D}=-1)(-1) \\ &= P(H)\cdot 1 + P(T)(-1) \\ &= p-(1-p) = 2p-1 \end{aligned}$$

This illustration could easily be extended into the earlier example of a two-toss series.

When one is talking about reaction time, the observed measurement (in the conventional "clock" sense) is already a number, so the mapping that represents the random variable associated with a reaction time is trivial, of course. Nevertheless, to be consistent when terms such as expectation are used, the reference must be to a random variable rather than the numerical event. Thus if $\mathbf{T}_i$ is the random variable representing t_i $(i=1,2)$, one should, to be precise, write $E(\mathbf{T}_1+\mathbf{T}_2)$ rather than $E(t_1+t_2)$ as the expectation of the sum. Now let us go back to develop the background of measures and probabilities in a little more detail. A measure m on a sigma field is a nonnegative function carrying the sets of the sigma field (specifically for our purposes, a Borel field $\mathfrak{B}$) into the positive reals, having the additional properties of being sigma additive, which is defined as

$$m\left(\sum_{i=1}^{\infty} b_i\right) = \sum_{i=1}^{\infty} m(b_i) \quad \text{for } b_i \cap b_j = \varnothing, \quad i \neq j$$

where each $b_i \in \mathfrak{B}$, and as mapping the empty set $\varnothing$ into the number 0. This immediately leads to the measure space $\langle S, \mathfrak{B}, m \rangle$, and if the total measure is 1, that is $m(W)=1$, then it is specifically a probability space. With $m=P$ we have generated $\langle W, \mathfrak{B}, P \rangle$. Sets that have measure 0 are said to be *null sets* (with respect to the measure m), and relations or definitions that are valid on

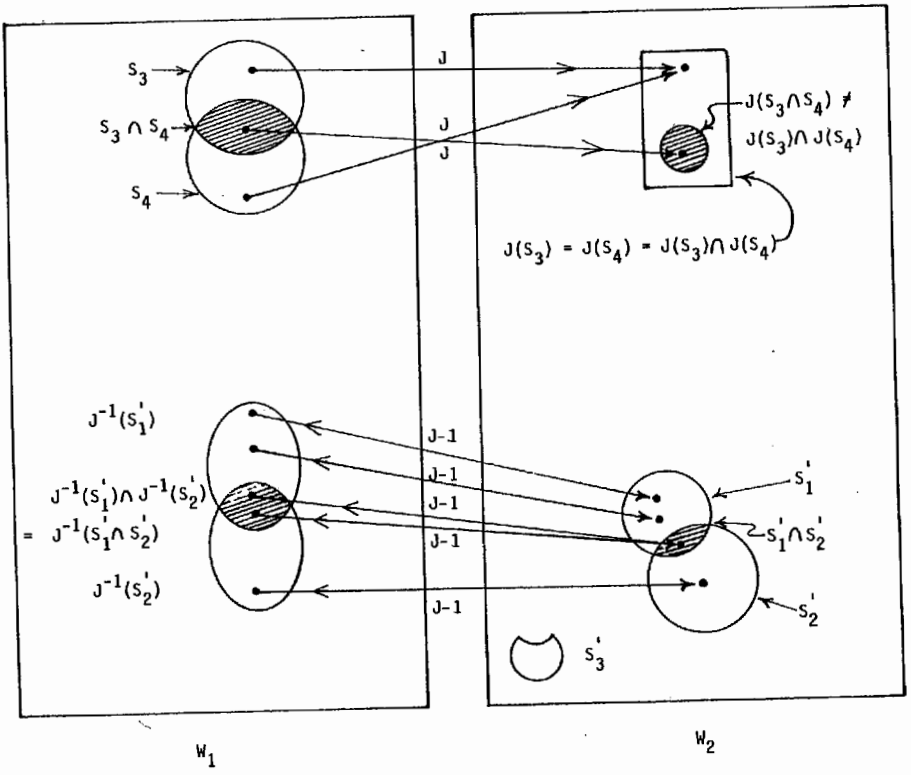


Fig. 14.1. Preservation of set operations in the inverse mapping and lack thereof in the original mappings.

all sets except those having measure 0 are said to be "valid almost everywhere."

Figure 14.1 shows a mapping J from one set of points to another. Notice that J itself does not preserve set relations because $J(S_3 \cap S_4) \neq J(S_3) \cap J(S_4)$ (top of figure). Note in fact that the intersection $S_3 \cap S_4$ is mapped onto the set of points represented by the circle. We can see in this particular case that J must be many-to-one because $J(S_3) = J(S_4) = J(S_3) \cap J(S_4)$; that is, each subset S_3 and S_4 is mapped to exactly the same set of points, represented by the rectangle. Next, as expected, the inverse mapping J^{-1} does preserve set operations, as indicated by $J^{-1}(S_1' \cap S_2') = J^{-1}(S_1') \cap J^{-1}(S_2')$ shown in a different mapping at the bottom of Fig. 14.1. Also here, the function J is not "onto" since the point set S_3' (at the bottom right half of the figure) has no inverse image [i.e., $W_2 \neq J(W_1)$]. No measures are shown on either set of points.

Now let M_i = model i and M_i = a class of models of which M_i is a member. The primary goal, of course, in investigating model implications and equivalence is to establish that if model M_1 holds, then model M_2 holds; more generally, that for any $\mathrm{M}_1 \in M_1$, there exists a model M_2 such that $\mathrm{M}_2 \in M_2$ that gives the same probability measure on all the corresponding possible events. Then the class M_1 is said to imply the class M_2. We will focus on model implication because model class implication results follow trivially. If the implication goes both ways, then equivalence of the models and model classes occurs. For this purpose, we need to amalgamate the probability measure and the Borel function. Let us go over the conditions that produce implication or equivalence and then briefly discuss them. The overriding consideration is that for every event with nonzero probability in M_1 there must be a corresponding event with the same probability in M_2 and that there be no events in M_2 with nonzero probability that have no correspondent events in M_1.

First, as indicated in Fig. 14.1, the mapping (function) J is from W_1 *into* W_2 and is single-valued, having exactly one image point in W_2 for *every* point in W_1. We shall see below that, in the case of one-way implication, J need not, in general, be *onto* W_2; that is, some points of W_2 may not be an image of any points in W_1. So far, we are just employing the usual definition of a function or mapping.

Secondly, the function J will usually be a measurable or Borel function in the sense established above. However, this is not a necessary condition for all model classes.

Third, it is further required that every event (Borel set) in $\mathfrak{B}_1$ having nonzero probability be an inverse image of some event (Borel set) in $\mathfrak{B}_2$; for if not, there would be no way to establish a probability on some event in M_2 that was equal to a probability on the given event in model M_1. It should be appreciated that simply making sure that every event in $\mathfrak{B}_1$ maps into an event in $\mathfrak{B}_2$ is not sufficient for this purpose, because that $\mathfrak{B}_1$ event might be one of other $\mathfrak{B}_1$ events that map into the same $\mathfrak{B}_2$ event. The inverse image of the $\mathfrak{B}_2$ event would then be a collection of $\mathfrak{B}_1$ sets whose union would itself be a $\mathfrak{B}_1$ set with its own probability, and it would then be impossible to establish a probability-equivalent event in $\mathfrak{B}_2$ corresponding to a particular event in $\mathfrak{B}_1$.

A fourth condition is the requirement that the probability associated with M_2 be such that the probability P_2 on any event in $\mathfrak{B}_2$ be *equal* to the probability P_1 on the inverse image of the $\mathfrak{B}_2$ event. Thus, $P_2(b_2) = P_1[J^{-1}(b_2)] = P_1(b_1)$, where b_1 maps onto b_2.

A fifth supplementary condition on model classes that we view as important for scientific as opposed to purely mathematical purposes is that the implied model class be in some sense at least as general as the implying model class. In most cases, the satisfaction of this condition will be obvious, but we will discuss it further below.

If, for any model M_1 in M_1 a model M_2 exists in M_2 that is implied by M_1, we say the class M_1 implies the class M_2. When M_1 implies M_2 and vice versa, M_1 and M_2 are *equivalent*. When M_1 implies M_2 and M_2 implies M_1, the two classes are equivalent.

Now let us look at the reasons for the above conditions, before extending

the discussion to equivalence. Why is the mapping J_1 not necessarily onto W_2? The reason involves the assignment of probability 0 to one or more sets in $\mathfrak{B}_2$. As an example, consider sets $W_1=\{0,1\}$, $W_2=\{0,1,2\}$, and the associated finite fields $\mathfrak{B}_1=\{\{0\},\{1\},\{0,1\},\varnothing\}$ and $\mathfrak{B}_2=\{\{0\},\{1\},\{2\},\{0,1\},\{0,2\},\{1,2\},\{0,1,2\},\varnothing\}$. Let $J\colon 0\to 0, 1\to 1, \varnothing\to 2$ (see Loève 1955: 62–63). Then J is a Borel (i.e., measurable) function since $J^{-1}\colon W_2\to W_1$ is as follows:

$$
\begin{aligned}
W_2 &\to W_1\\
\{2\} &\to \varnothing\\
\varnothing &\to \varnothing\\
\{0\} &\to \{0\}\\
\{1\} &\to \{1\}\\
\{0,1\} &\to \{0,1\}
\end{aligned}
$$

Note that the other sets contained in $\mathfrak{B}_2$ have no inverse image (except $\varnothing$) and so satisfy the measurable function criteria vacuously.

All these (inverse image) sets are contained in $\mathfrak{B}_1$. Let $P_1(\{0\})=\frac{1}{2}$, $P_1(\{1\})=\frac{1}{4}$, $P_1(\{0,1\})=\frac{1}{4}$, $P_1(\varnothing)=0$, and let P_2 be such that $P_2(\{0\})=\frac{1}{2}$, $P_2(\{1\})=\frac{1}{4}$, $P_2(\{0,1\})=\frac{1}{4}$, and $P_2(\text{any other set})=0$. This setup satisfies the other conditions and makes M_2 give an identical probability distribution on the events as M_1; yet J was clearly not onto. It is certainly true, on the other hand, that if a set in $\mathfrak{B}_2$ has a nonzero probability, there must be an inverse image of that set with the same probability.

The second requirement, that the mapping J be a measurable function, should be pretty cogent since in general we would otherwise be trying to associate things that are events in one M_2 model with nonevents in M_1. However, in certain situations, this condition is unnecessary. An example connected with the third condition will be used to illustrate this point below.

The third condition is crucial for our purposes. In this connection, a mapping can be onto in the sense that every point (member) in W_2 has an inverse image in W_1 without necessarily producing a corresponding set in $\mathfrak{B}_2$ for every set in $\mathfrak{B}_1$. An example is:

$W_1=\{0,1,2\}$; $\mathfrak{B}_1=\{\{0\},\{1\},\{2\},\{0,1\},\{0,2\},\{1,2\},\{0,1,2\},\varnothing\}$

$W_2=\{0,1,2\}$; $\mathfrak{B}_2=\{\{1\},\{0,2\},\{0,1,2\},\varnothing\}$

Let $J\colon \varnothing\to\varnothing, 0\to 0, 1\to 1, 2\to 2$ to satisfy $W_1\to W_2$ being onto (and one-to-one), but there are $\mathfrak{B}_1$ sets, for example, $\{0,1\}$ which have no correspondent in $\mathfrak{B}_2$. Incidentally, observe that $\mathfrak{B}_2$, although a coarser field than $\mathfrak{B}_1$ (if $b\in\mathfrak{B}_2$, then $b\in\mathfrak{B}_1$), still obeys the restriction that it is closed under all set operations on its constituent sets.

In any case, in this situation, even though J^{-1} is measurable, by mapping sets in $\mathfrak{B}_2$ to sets in $\mathfrak{B}_1$, it is impossible for $M_1\to M_2$ when, say, $P_1(\{2\})>0$, for there is no comparable event in M_2. Moreover, one cannot necessarily circumvent the difficulty by formulating a set mapping that is many-one from $\mathfrak{B}_1$ onto $\mathfrak{B}_2$. For example, suppose that $\{0,2\}$ as well as $\{0,1\}$ in $\mathfrak{B}_1$ were

mapped onto $\{0,2\}$; that is, let $J\colon 0\to 0, 1\to 2, 2\to 2$. Then, the inverse image of $\{0,2\}$ in $\mathfrak{B}_2$ is the set $\{0,1,2\}$, providing no structure whereby M_2 can emulate the probabilities on events such as $\{0,2\}$ in M_1. This example therefore does not satisfy the third condition.

Now we indicate how measurability of the function J may sometimes be circumvented. Consider again the example of the third condition, but now map in the reverse direction. That is, we attempt to form a model M_1 that is identical to M_2. Suppose the component sets of $\mathfrak{B}_2$ all have nonzero probability except $\varnothing$. Let $J\colon W_2\to W_1$ by $J\colon \varnothing\to\varnothing, 0\to 0, 1\to 1$. Then the $\mathfrak{B}_1$ sets $\{1\}$, $\{0,2\}$, $\{0,1,2\}$, $\varnothing$ have the obvious inverse images and can be given probabilities equal to those of their inverse image events. Further, none of the other components of $\mathfrak{B}_1$ have inverse image events in $\mathfrak{B}_2$; yet they can simply be assigned probability 0 to yield an equivalent M_1 model.

The fourth condition sets up the probability measure that is associated with the Borel function and yields a set of probability equivalent events in model M_2 that mimic the events and their probabilities in model M_1. Why are we compelled to employ the inverse mapping J^{-1} to establish that any $\mathfrak{B}_2$ set (event) has a probability that agrees with the probability space of M_1? That is, we wish to find the b_2 terms and P_2 so that $P_2(b_2)=P_1[J_1^{-1}(b_2)]=P_1(b_1)$ for each b_1. Why cannot we use the mapping J itself? First of all, it is the event in M_2 that must satisfy the probabilities and related events of M_1, so we want to make M_2's probabilities mimic those of its inverse images. A related point is that, as we saw, several events might map to the same M_2 event. Secondly, the original function J itself need not preserve set operations as we saw. Suppose, for instance, that b_1 and b_1', both belonging to $\mathfrak{B}_1$, map into the same Borel set in $\mathfrak{B}_2$. Then $P_1(b_1\cap b_1')$ has a well-defined meaning in M_1, but $P_2[J(b_1\cap b_1')]$ may be completely unrelated to the individual measures on b_1 and b_1' because the latter will not usually equal $P_2[J(b_1)\cap J(b_1')]$. Put another way, $b_1\cap b_1'$ is a distinct event in $\mathfrak{B}_1$ and may map into something in $\mathfrak{B}_2$ quite different from the separate mappings of b_1 and b_1'. On the contrary, $J^{-1}[J(b_1\cap b_1')]$ maps this image of $b_1\cap b_1'$ back to itself and hence provides the basis for assigning a probability to $J(b_1\cap b_1')$ that is consonant with the probability of $b_1\cap b_1'$ as established in model M_1.

The fifth condition, that in some sense M_2 be at least as general as M_1, is not fully formalized, but in actual cases it will ordinarily be evident whether this holds. For instance, consider the parallel and serial model classes based on exponential intercompletion times (Chapter 4). Although the sets of points and, indeed, the event spaces are of equal size, it is, we may recall, the distributions on these sets of events included in the sigma space that are different. For the present example, let us return to working with the intercompletion times themselves. Then, when $n=2$, the set $W_s=W_p=\mathbf{R}^+\times\mathbf{R}^+=\mathbf{R}_2^+$ is the two-dimensional space of positive real numbers and so the point sets are the same for the serial and parallel models. Clearly, the events (sets) of the generated Borel (sigma) fields consist of intervals and set operations (and limits on a countably infinite number of such operations) on those intervals

of real numbers in $\mathbf{R}_2^+$, and they are obviously equally general, being actually identical. However, it is the probability distributions on these spaces that differ, since the parallel model predicts that the distribution on the minimum completion time, conditional on which position finishes first, will be the same

$$(v_{a1}+v_{b1})\exp[-(v_{a1}+v_{b1})t]$$

for the two positions, whereas the serial model allows them to be distinct:

$$u_{a1}\exp(-u_{a1}t) \neq u_{b1}\exp(-u_{b1}t)$$

The serial model class is more general in this particular context, and it is reasonable to say that it is implied by the parallel model class but not vice versa, because the serial class may be satisfied when the parallel class is not. Again, a restriction on the parameters of the serial model, as we know from our earlier work, reduces the generality of the serial class to a special case that is equivalent to the parallel model (the constraint is $u_{a1}=u_{b1}$).

In some cases, the more general model class M_2 may simply possess an event space that is a refinement ($\mathcal{B}_1$ is coarser than $\mathcal{B}_2$, i.e., if $b\in\mathcal{B}_1$, then $b\in\mathcal{B}_2$) of the M_1 event space. In the example above, where $W_1=\{0,1,2\}=W_2$, $\mathcal{B}_1=\{\{0\},\{1\},\{2\},\{0,1\},\{0,2\},\{1,2\},\{0,1,2\},\varnothing\}$, $\mathcal{B}_2=\{\{1\},\{0,2\},\{0,1,2\},\varnothing\}$, $\mathcal{B}_1$ is a refinement of $\mathcal{B}_2$. Therefore, if all the events of $\mathcal{B}_1$ are of nonzero probability, it would be reasonable to attempt to achieve the implication $M_2 \rightarrow M_1$ but not $M_1 \rightarrow M_2$.

Finally, consider an interesting example where a simple ordering in terms of generality may not be possible. Suppose a model class M_1 yields probability distributions on a series of correct and error responses, whereas a more general model class M_2 might give that plus probability distributions on RTs. In a sense, M_2 seems more general than M_1 because M_2 attempts to explain more than M_1, and in fact, contains the information of the latter. However, intuitively we might expect falsification of a more general class of models to imply falsification of the less general, but that need not happen here. For suppose the RT predictions of M_2 are not supported by a set of data, whereas both classes correctly predict probability correct and related statistics. Then M_1 is verified whereas M_2 is falsified. On the other hand, when M_2 is correct, it certainly explains more of the "universe." There is a way of bringing this kind of consideration into a more formal statement of generality: Require that an ordering of generality with respect to implication and equivalence relations be based on an identical set of atomic events in the probability space. This would effectively rule out the present example, where the model space gets bigger by including an additional event space (technically, the Cartesian product is taken of a probability correct × RT on each trial in the present example). This type of concern will probably lead to the generality ordering being a partial order, since many model classes may not be comparable. The classes of equivalence classes composed of equivalent models would then lie in a partial order (i.e., a binary relation on the set of equivalence classes relation that is transitive, reflexive, and antisymmetric). In this context, we could partially order the classes of models based on exponential intercompletion

times investigated in Chapter 5, using $\gtrsim$ as the order symbol: compound serial $\gtrsim$ regular serial $\gtrsim$ regular parallel; compound serial $\gtrsim$ compound parallel $\gtrsim$ regular parallel; but regular serial and compound parallel are not ordered, that is, are "noncomparable," because neither is more general than the other (see Fig. 5.1). Of course, most noncomparable model classes are less interesting, being of the apples vs. oranges variety (e.g., models of schizophrenia vs. models of brightness perception).

Equivalence of parallel and serial models

It is now time to specialize the development and apply it to the general parallel–serial equivalence problem, again with the proviso that realistic aspects of material systems that can aid in producing discriminable features of these processes are put aside or ignored.

Until the mid-1970s parallel–serial equivalence results were largely confined to particular families of distributions (e.g., exponential or Weibull). Townsend (1976b) and Townsend and Romei (1974) produced equivalence mappings between general serial models and within-stage-independent parallel models; results that have been alluded to elsewhere in the book (in particular, Chapter 4). Anderson (1976) exhibited the form that general parallel–serial equivalence functions would take in the case of $n=2$ elements. Vorberg in 1977 presented equivalence theorems at the level of generality addressed in this chapter, although our treatment differs in certain respects.

Let the set of points relevant to the parallel model be designated

$$W_p=\{(\tau_1,\tau_2,\ldots,\tau_n;a_1,a_2,\ldots,a_n)\mid \tau_i\in\mathbf{R}^+,\quad 1\leqslant i\leqslant n;$$
$$\tau_1\leqslant\tau_2\leqslant\cdots\leqslant\tau_n,\quad (a_1,a_2,\ldots,a_n)\in\text{permutation}(n)\}$$

where τ_i is the (total completion) time at which the ith element to be finished is completed, irrespective of the identity or serial position of the element, and a_i is the serial position of the element completed ith. Recall that τ_i is also the actual processing time of element i. Similarly, $\mathbf{T}_i$ is the random variable associated with the occurrence τ_i, although we really do not require it in the following discussion.

Similarly, the set of points employed for the serial model is

$$W_s=\{(t_1,t_2,\ldots,t_n;a_1,a_2,\ldots,a_n)\mid t_i\in\mathbf{R}^+,\ 1\leqslant i\leqslant n;$$
$$(a_1,a_2,\ldots,a_n)\in\text{permutation}(n)\}$$

The t_i are values of the intercompletion time random variable $\mathbf{T}_i$ and are, of course, also the actual serial processing times. The a_i and other notation have the same meaning as in the parallel case. This notation is identical to that employed in the section "Generalization to n elements" of Chapter 4. Note that when $n=2$, this notation differs from our usual where a and b represented the two serial positions and u_{ai} and u_{bi} ($i=1,2$) were the rates on positions a or b during stage i. We now let a_i be a number between 1 and n, so $(a_1,\ldots,a_n)$ is a permutation of the numbers 1 through n. It is not particularly

important for present goals whether a is given as the serial position of the element, as we have done, or as the identity of the element, because none of the properties of parallel or serial models that can potentially discriminate them are employed in the following. Indeed, we could employ a double index indicating both serial position and identity, for all that it would matter in the equivalence results. The only slight preference for serial position here, if one is to be chosen, is that we will still have the concept of p as the probability of an order in the serial models. An order selected on the basis of serial position will, of course, induce an order on the identities of the elements. However, as we indicated earlier, it seems more natural to suppose that a serial system can (possibly) select an order based on the serial position than on identity, a priori, so we continue to use that here. We are now in a position to state the general definitions of parallel and serial models.

Definition 14.1: A *model* of a *parallel* system is defined by $\mathrm{M}_p = \langle W_p, \mathfrak{B}_p, P_p \rangle$, where $\mathfrak{B}_p$ is the Borel field associated with W_p (and therefore the completion = actual processing times τ) and P_p is some particular probability measure on $\mathfrak{B}_p$. For every "point" w belonging to W_p, the probability density function assigned by M_p is given by

$$g_p(w) = g(\tau_1, \tau_2, \ldots, \tau_n; a_1, a_2, \ldots, a_n)$$

with

$$w = (\tau_1, \tau_2, \ldots, \tau_n; a_1, a_2, \ldots, a_n)$$

and the actual dimensional values are τ_1 on dimension a_1, τ_2 on dimension a_2, and so on. We denote the general *parallel model class* as $M_p = \{\mathrm{M}_p\}$. □

And next, the serial model definition:

Definition 14.2: A *model* of a *serial* system is defined by $\mathrm{M}_s = \langle W_s, \mathfrak{B}_s, P_s \rangle$, where $\mathfrak{B}_s$ is the Borel field on W_s (and therefore on the intercompletion = actual processing times t). The probability measure P_s has our typical structure with a probability distribution on the order of processing and then a distribution on actual processing times that is conditioned on that order. Thus, $P(a_1, a_2, \ldots, a_n)$ gives the probability of the order shown. The conditional probability density function is

$$f(t_1, t_2, \ldots, t_n \mid a_1, a_2, \ldots, a_n)$$

The joint probability function is then, naturally,

$$f_s(w) = P(a_1, a_2, \ldots, a_n) f(t_1, t_2, \ldots, t_n \mid a_1, a_2, \ldots, a_n)$$

and finally, the general *serial model class* is $M_s = \{\mathrm{M}_s\}$. □

Our plan of attack will be to provide a mapping between the parallel and serial spaces that is one-to-one and onto and to show that for any model of one type a model of the other type exists that is equivalent.

Proposition 14.1: The serial and parallel model classes defined in Definitions 14.1 and 14.2 are equivalent.

Proof: (i) M_s implies M_p: Consider an arbitrary serial model $\mathrm{M}_s = \langle W_s, \mathfrak{B}_s, P_s \rangle$. Let J_{sp} be the mapping from the serial to the parallel space, $W_s \to W_p$, and in fact suppose that

$$J_{sp}(t_1, t_2, \ldots, t_n) = (\tau_1, \tau_2, \ldots, \tau_n) = \left(t_1, t_1 + t_2, \ldots, \sum_{j=1}^{n} t_j\right)$$

where $\tau_i = \sum_{j=1}^{i} t_j$, $1 \leqslant i \leqslant n$. Now we have to prove that a probability function (which in general can be a density or mass at any point in the same spaces) can be provided in the parallel space that is equivalent to the serial probability function on the inverse image of J_{sp} in the serial space. This will show that the serial model implies the parallel model.

The parallel probability function, it may be recalled, is

$$g(\tau_1, \tau_2, \ldots, \tau_n; a_1, a_2, \ldots, a_n)$$

From our above definition of J_{sp} it follows that the inverse map, $J_{sp}^{-1} = J_{ps}: W_p \to W_s$, is given by

$$J_{ps}(\tau_1, \tau_2, \ldots, \tau_n) = (\tau_1, \tau_2 - \tau_1, \ldots, \tau_n - \tau_{n-1})$$

so that

$$g(\tau_1, \tau_2, \ldots, \tau_n; a_1, a_2, \ldots, a_n) = g\left(t_1, t_1 + t_2, \ldots, \sum_{1}^{n} t_i; a_1, a_2, \ldots, a_n\right)$$

We now set this parallel function equal to

$$\begin{aligned} g(\tau_1, \tau_2, \ldots, \tau_n; a_1, \ldots, a_n) &= pf[J_{sp}^{-1}(\tau_1, \tau_2, \ldots, \tau_n)] \\ &= pf[J_{ps}(\tau_1, \tau_2, \ldots, \tau_n)] \\ &= pf(t_1, t_2, \ldots, t_n \mid a_1, a_2, \ldots, a_n) \end{aligned}$$

The parallel probability function g is thus arbitrarily set equal to its serial correspondent for every point $(t_1, t_2, \ldots, t_n)$ in the serial space. Because the serial model was arbitrary, a corresponding parallel model exists for every serial model in M_s so that M_s implies M_p.

(ii) M_p implies M_s: To move in the reverse direction and prove that the parallel class implies the serial class of models, it is convenient to first find what p must be in terms of the parallel process,

$$p = \int_0^\infty \int_{\tau_1}^\infty \cdots \int_{\tau_{n-1}}^\infty g(\tau_1', \tau_2', \ldots, \tau_n'; a_1, a_2, \ldots, a_n)\, d\tau_n'\, d\tau_{n-1}' \cdots d\tau_1'$$

so p correctly gives the probability that the elements are finished in the order (according to serial position) $a_1, a_2, \ldots, a_n$. Next, the serial conditional probability function must be set so that the mapping from its space back to the parallel space preserves the probability structure. That is,

$$f(t_1, t_2, \ldots, t_n \mid a_1, a_2, \ldots, a_n)$$
$$= \frac{1}{p} g\left(t_1, t_1+t_2, \ldots, \sum_{i=1}^{n} t_i; a_1, a_2, \ldots, a_n\right)$$
$$= \frac{1}{p} g(\tau_1, \tau_2, \ldots, \tau_n; a_1, a_2, \ldots, a_n)$$

where p represents the integral of the joint probability function (the solution of p given just above). Because we can make the conditional function f anything we like at any actual serial processing times $t_1, t_2, \ldots, t_n$, the above equation can certainly be satisfied in such a way as to correspond in a one-to-one fashion to the parallel actual processing times $\tau_1, \tau_2, \ldots, \tau_n$, as the above expressions demonstrate. Again the parallel model was arbitrary, so clearly M_p implies M_s. □

As pointed out earlier, these mappings do not prevent untoward situations that arise when realistic systems are investigated. For instance, to mimic a serial model with the property that the first processing time depends on the later order of processing (according to position) a parallel model may have to foretell the future. Conversely, assume that the identity of each element rather than the position is considered and that the goal of processing is to establish identity of the elements. Then parallel times can depend on later finishing order (according to identity) but serial times cannot without being prescient.

Two examples may help to elucidate the ideas in the foregoing treatment. Let us construct a parallel model equivalent to a serial exponential model that has distinct rates on the first processing time that depend on the position processed. Let $n=2$, and suppose a and b are the two pertinent serial positions so that a_1 can be a or b and a_2 can also be a or b. Thus, the relevant rates are

Stage 1: $u_1 = u_{a1}$ when $a_1 = a$
$= u_{b1}$ when $a_1 = b$

Stage 2: $u_2 = u_{a2}$ when $a_2 = a$
$= u_{b2}$ when $a_2 = b$

This is the notation we have used throughout the book when $n=2$. Recall that in Chapter 4 we saw that no independent parallel exponential model is equivalent to this serial model. In fact, we know that no within-stage-independent parallel model is identical to this model (see Chapter 4).

We now desire the mapping from the serial to the parallel space, $\tau_i = \sum_{j=1}^{i} t_j$, $1 \leqslant i \leqslant n$, and then

$$g(\tau_1, \tau_2; a_1=a, a_2=b) = pf(t_1, t_2 \mid a_1=a, a_2=b)$$
$$= pu_{a1} \exp(-u_{a1} t_1) u_{b2} \exp(-u_{b2} t_2)$$
$$= pu_{a1} \exp(-u_{a1} \tau_1) u_{b2} \exp[-u_{b2}(\tau_2 - \tau_1)]$$

and next

$$g(\tau_1, \tau_2; a_1=b, a_2=a) = (1-p) f(t_1, t_2 \mid a_1=b, a_2=a)$$
$$= (1-p) u_{b1} \exp(-u_{b1} t_1) u_{a2} \exp(-u_{a2} t_2)$$
$$= (1-p) u_{b1} \exp(-u_{b1} \tau_1) u_{a2} \exp[-u_{b2}(\tau_2 - \tau_1)]$$
$$t_1, t_2 \geqslant 0$$

The reader will observe that $u_{a1} \neq u_{b1}$ (in general) has been allowed even though in the parallel case it implies a dependence on future events. Thus no processing interpretation is presently available for the above resulting parallel model. In fact, among other problems, it does not obey the principle of stochastic simultaneity for parallel processing given in the next chapter. Also note that it is not exponential in the usual sense; for instance, p is not in general related to the rates as in true parallel exponential models.

As a final example, consider the parallel model based on the following two independent Rayleigh distributions:

$$g(\tau_1, \tau_2; a_1=a, a_2=b) = g(\tau_{a1}, \tau_{b2}) = g(\tau_a, \tau_b)$$
$$= a\tau_a \exp\left(-\frac{a\tau_a^2}{2}\right) b\tau_b \exp\left(-\frac{b\tau_b^2}{2}\right), \quad \tau_a, \tau_b \geqslant 0$$

Being independent, the expression does not depend on whether a or b is completed first. The Rayleigh distribution is not an implausible model for latencies, possessing as it does an increasing hazard function and being skewed to the right. It has a mean of $E(\mathbf{T}) = \pi/2a$ and variance $\text{var}(\mathbf{T}) = (2-\pi/2)(1/a)$ and is a special case of the Weibull and χ^2 distributions. Solving for the serial parameters we find first that

$$p = \int_0^\infty a\tau_a \exp\left(-\frac{a\tau_a^2}{2}\right) \exp\left(-\frac{b\tau_a^2}{2}\right) d\tau_a$$
$$= \frac{a}{a+b} \int_0^\infty (a+b)\tau_a \exp[-\tfrac{1}{2}(a+b)\tau_a^2]\, d\tau_a = \frac{a}{a+b}$$

so the interesting outcome appears that p is (as in the exponential case) equal to the relative a-to-b parameter values.

As for the serial processing time density:

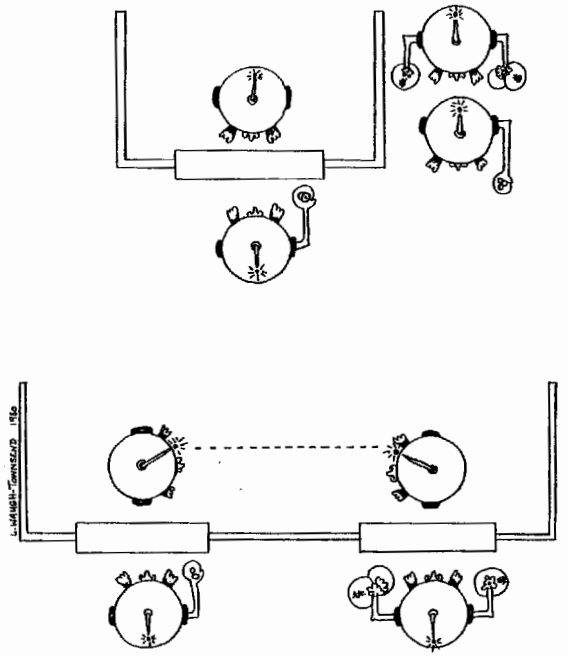
$$f(t_1, t_2 \mid a_1=a, a_2=b)$$
$$= \frac{1}{p} a\tau_a \exp\left(-\frac{a\tau_a^2}{2}\right) b\tau_b \exp\left(-\frac{b\tau_b^2}{2}\right)$$
$$= \frac{a+b}{a} at_{a1} \exp\left(-\frac{at_{a1}^2}{2}\right) b(t_{a1}+t_{b2}) \exp\left[-\frac{b(t_{a1}+t_{b2})^2}{2}\right]$$
$$= (a+b)\, bt_{a1}(t_{a1}+t_{b2}) \exp\left[-\frac{(a+b)t_{a1}^2 + 2bt_{a1}t_{b2} + bt_{b2}^2}{2}\right]$$

By the same token,

$$
\begin{aligned}
&f(t_1, t_2 \mid a_1 = b, a_2 = a) \\
&= \frac{1}{(1-p)}\, b\tau_b \exp\left(-\frac{b\tau_b^2}{2}\right) a\tau_a \exp\left(-\frac{a\tau_a^2}{2}\right) \\
&= \frac{(a+b)}{b}\, bt_{b1} \exp\left(-\frac{bt_{b1}^2}{2}\right) a(t_{b1}+t_{a2}) \exp\left[-\frac{a(t_{b1}+t_{a2})^2}{2}\right] \\
&= (a+b)\, at_{b1}(t_{b1}+t_{a2}) \exp\left[-\frac{(a+b)t_{b1}^2 + 2at_{b1}t_{a2} + at_{a2}^2}{2}\right]
\end{aligned}
$$

and the example is complete. With no further structure imposed on the parallel distributions, this example does not assault one's intuitions concerning realistic systems, unlike the first example.

In the succeeding chapter we again take up aspects of parallel and serial processing that help to differentiate and potentially test between them.



Commercial life on Jupiter exhibits a constraining aspect of *realistic serial systems,* as shown in the top part of the figure. The Jovian bank teller here is unable to see that a customer down the line (around the corner) is extraordinarily wealthy and that he should perhaps speed up his service to the interposed customers. Contrastingly, the two tellers operating in parallel below are shown in communication with one another, so the fact that one teller is working with an affluent customer can affect the behavior of the other teller. The critical idea here is that a serial processor should in principle not be able to use (know, etc.) any information about an item that is processed later. A parallel processor, because it works on everything simultaneously, can evidence such interitem effects. For instance, the identity of one item can then legitimately influence the processing rate of another item. (See property IIC in the text.)

15 *A general discussion of equivalent and nonequivalent properties of serial and parallel systems and their models*

Introduction

The strategy here will be to provide a characterization of stochastic serial and parallel *systems,* in terms of properties that appear to be reasonable for such systems that we might actually expect to find working in real time. We shall

then attempt to tie these properties into the *models* intended to describe the real systems. Some qualitative and some quantitative statements will be made with regard to distinctive aspects of the models. When none of the distinctive properties hold, then, as might be anticipated by this stage in our developments, parallel–serial equivalence will occur (see the preceding chapter).

This general procedure is in the spirit of an earlier treatise on serial and parallel equivalence and diversity in models based on exponential intercompletion times (Townsend 1976a). There are two aspects of this approach that deserve comment. The first is that the emphasis on the modeling of real systems bestows the saving grace of potential testability. The reason is that the realistic properties of such systems are responsible for the parallel–serial diversity, when combined with the respective simultaneity or sequentiality of the two types of processing. When the *only* properties considered are simultaneous vs. sequential time variables with distributions imposed on them, there is little structure left to lend some distinguishability to the associated models. In this way, the approach is not completely unlike that of earlier investigators who associated extra properties with parallel or serial processing (e.g., independence of total completion times, or perhaps limited capacity with parallel processing). On the other hand, we do feel that the aspects of parallel or serial processing that we are currently associating with real functioning systems are more integral and fundamental than has been the case in previous characterizations. This is not to say that all fundamental parallel–serial differences are encapsulated in this chapter. A number of others have been developed elsewhere in the book. The present chapter pinpoints several distinctions not treated heretofore or not previously given sufficient attention.

Conceptually related to this listing of presumably reasonable properties of the systems is the treatment of questions of equivalence (indeed, of any predictions) as being relative to the experimental or real-time context in which the systems are thought to be operable. This is a little different from the usual approach in which a given experimental milieu or specific paradigm is assumed and the pertinent models presented that apply to that specific situation. Questions of possible equivalence or mimicking are then sometimes posed at that stage. In some instances this procedure suffices very well. However, it can also be interesting to postulate the important characteristics of a system and then to see what happens when that system is placed in a particular experimental or real-life circumstance. For instance, it has been shown to be advantageous to employ the assumption that when elements are compared with one another, matching elements may be compared at a different speed than mismatching elements (Chapter 13; Townsend 1976a; Snodgrass & Townsend 1980). To be sure, this increased generality sometimes exacerbates equivalence problems, as when self-terminating models can thereby mimic exhaustive predictions (Townsend & Roos 1973) or vice versa (Townsend 1974b).

One aspect of greater generality in this chapter will be the concept of a state of processing for each element. This harks back to some earlier papers

(Townsend 1969; Townsend 1972: 169 ff.) in which the input was conceived of as a vector, itself composed of subvectors. For example, the constituents of the original vector might be letters, while the components of each letter vector might be features. The output was also a vector function of time that gave the state of processing of each of the elements at any given time t. However, the emphasis was always on the finishing time of the elements themselves rather than on the evolution of the processing states of the elements up until they are completed. In fact, the most examined distributions, those based on exponential processing at the element level, are not compatible with more than a bivariate set of states, *uncompleted* and *completed*. In any case, we shall see that when the set of states includes more than two states, the resultant possibility of partial processing yields a particular line of potential for parallel–serial testability.

In the most general cases, the state might consist of the names of the components of the elements, and it might be that the components could finish in more than one order from trial to trial, in analogy to what happens with the more macroscopic elements or their serial positions. Sometimes it might not matter what the identity of the components is, or they may have no names, in which case the state of processing on an element might simply be related to the number of components completed at a given point in time. The latter instance would correspond to the counting process that is the discrete state counterpart of a waiting-time process, as one sees, for example, with the Poisson counting process and the gamma waiting-time process. The idea of a *global state* of the system will also be useful. The global state simply logs the complete state description of all the individual elements at any point in time. For instance, the names of all the completed components of each element might be given throughout $t \geqslant 0$ in a certain system.

The list of properties below will include some of the aspects that differentiate serial and parallel processing as well as some of the major features shared by the two, although the ensuing discussion will emphasize diversification. Although the various properties seem to reflect different facets, we are not in a position to assert their logical independence at this point; some of them are palpably related. After the set of properties is surveyed, an important principle of mathematical models and theory testing will be considered – a principle that is rather obvious, but not always fully appreciated. The name we give this property is the *principle of correspondent change*.

The properties of the parallel and serial systems that are alike will be put down first, followed by differentiating aspects. The properties concerning processing times will usually be stated as such, with consequences for the stochastic distribution on those times to be taken as implied.

There are several directions that could be taken with regard to the statement and discussion of the properties. One approach would be to define a general probability space with all the potential dependencies on the various properties explicitly entered into the notation. However, experience with the earlier case of pairs of objects being compared exponentially, with fewer dis-

tinctions than we are prepared to make here, suggests that the goal of readability and general comprehensibility might well suffer due to the great complexity of notation required for a complete description. We are therefore going to try to simply establish necessary notation as we go along, leaving out structure and associated notation that is irrelevant to a particular point.

Thankfully, some of the distinctive aspects of serial and parallel processing allow for testability, although others appear to be obscured and rendered unusable experimentally in the generality of the probability spaces and the scant observability in the usual laboratory setting. It is apropos to mention in this context that observability and consequent testability might well be better when the present model concepts are applied in certain other cognitive areas of psychology. Any paradigm in which the whole stochastic process (or a realization thereof) is run through every single trial is likely to present difficulties in identification of underlying mechanisms. Many of the perceptual and simple memory tasks emphasized in this book are of this variety. There are other areas, such as problem solving, sentence comprehension, or extended visual search, where a greater part of the stochastic process may be laid out before one, as it were. An example of past research where the nature of the problem has permitted more observability is learning theory, where each trial is only one small part of the stochastic process presumed to explain the observed behavior, thus enhancing model testability.

Natural properties of parallel and serial systems and their models

I. Properties held in common

IA. The processing times and the order of completion of the elements can depend on the serial positions of the element being processed as well as the serial positions of the completed elements.

IB. The processing time on a given element can depend on the identity of that element as well as on the identities of the already completed elements and their processing durations.

IC. The order of completion of the elements can depend, in an evolving way, upon the identity and processing times of previously completed elements.

ID. The probability that the global state of the system changes during any finite Δt is nonzero, no matter how small Δt is (but $\Delta t > 0$).

II. Divergent properties of serial and parallel systems

IIA. Description of the global state of the system

Serial: The global state of the system is always given by i $(0 \leqslant i \leqslant n)$ completed elements, the partially completed status of one element [e.g., the completed components of at most one uncompleted element (up until $i=n$)],

and $n-i-1$ uncompleted elements, together with the identity, serial position information, order, and associated times of the completed elements.

Parallel: The global state of the system is always given by i completed elements, and the partially completed status of all other $n-i$ elements, for $0 \leqslant i \leqslant n$, together with the identity, serial position information, order, and associated processing times of the completed items.

IIB. Change of state of individual elements

Serial: The probability that any element other than the one currently undergoing processing changes state during an interval Δt, given that the current element is unfinished during that interval, is 0.

Parallel: The probability that any uncompleted element can change its state during any interval Δt is always nonzero, no matter what other element is "given" as being uncompleted.

IIC. Dependence of processing time of an element on the identity of uncompleted elements

Serial: The processing time of an element cannot depend on the identity of the elements completed after it or the processing order expressed in terms of identity.

Parallel: The processing time of an element can depend on the identity of elements completed after it and on the processing order expressed in terms of identity.

IID. Dependence of processing times on a preordained order

Serial: It is possible to design and build a serial system so that a processing order based on the serial positions (but not the identities) of the elements is preselected and then to make the processing times depend on the particular preordained order on any given trial.

Parallel: A parallel system cannot be designed and built so that a processing order based on the serial positions can be preselected; therefore, the processing times cannot depend on such a preordained order.

Principle of correspondent change

A. Empirical changes in the environment or stimulating situation should be reflected in a nonvacuous theory or model by corresponding

changes or invariances in the model or theory. Such changes or invariances are predictable and in consonance in the correct model or theory but are not predictable and are inharmonious with an incorrect model or theory.

B. For any given empirical milieu and for any given class of models, there will exist a set of subclasses where models are nondistinguishable within their subclass and in that specific milieu.

Two theories or models may, of course, make predictions at some level that are equivalent (e.g., the means for a given change in experimental conditions). However, to the extent that the two theoretical structures are not truly identical, or, in effect, at least not empirically identical, there must exist aspects of the models or theories that relate to different possibilities in the real world. It is obvious (and tautological) that scientists implicitly employ this principle on a day-to-day basis. However, it has been of value in our own research to emphasize its theoretical importance and occasionally subtle manifestations.

General discussion

Analogous properties: IA–D

Property IA is one we have dealt with at length throughout the book and so it requires little comment here. The idea is that either type of system may have the ability to employ information about completed elements in setting the rates (more generally, the probability distributions on processing times) on processing. Similarly and obviously, the processing time of an element can depend on its own serial position. With regard to order of completion, in a serial system the processing path distribution can in principle be set up beforehand in terms of permutations on the serial positions of the elements. In an alternative type of serial system a decision could be made as to which position to process first. Then, based upon that choice, a conditional probability distribution on the remaining positions determines the second position to be processed, and so on. In parallel processing, on the other hand, the probability distribution on processing times can be a function of the different positions, and this distribution can change depending on which positions are finished in what order, as the processing evolves. Thus, the distribution on order of completion is more intimately tied to the processing time distributions in parallel than in serial systems.

In an analogous fashion, an element's processing time can clearly depend in general on its own identity as well as the identities and times of previously finished elements. This is property IB. The processing time distributions in a parallel system can depend on the identity and processing times of previously finished elements in two major ways. First, the distribution can be altered as elements are completed, as has been illustrated many places in the book with exponential rates, the rates changing according to the history of processing.

Second, when an element is composed of more than a single component, the changes that take place in the status of all the uncompleted elements during a single intercompletion time can affect the processing distribution, in general, of any of the elements. (This type of effect was not considered in earlier chapters.) By way of contrast, serial systems can make their elements' time distributions depend on previously completed elements directly, as in the first type of parallel effect, but the second type of effect is impossible in serial systems.

Property IC is a bit more subtle with regard to parallel and serial systems, primarily because we have not discussed the serial system under such circumstances before. As in IB, the parallel system's orders of completion and the distribution on those orders come naturally out of the evolving events of processing.

Also in analogy to IB, but not heretofore discussed in this book, a serial system could make the *order* (path) of processing depend not only on the identity of previously finished elements but also on their processing durations. For example, a probability distribution could be conditioned on previous elements as well as their processing times and thereafter defined on the possible processing orders of the remaining elements. It is very important to note that a serial system operating in this fashion, with the order depending on previous processing times, cannot possess full flexibility with respect to preselecting orders of processing, and vice versa. Thus, Property IID (discussed below) says that, in general, a serial system can preordain the processing orders and make the time distributions be functions of these orders. But if the order in a given system "evolves" as the history of the process is being laid down, obviously this preordaining flexibility cannot be realized. It would be interesting to investigate such systems in more detail, particularly those with a mixture of historical time dependence and preselection, but this will have to be a future enterprise. The ensuing discussion will not further consider serial systems whose processing *order* is based on the history of processing times.

Property ID is just a way of saying that the processing never completely stops, in either a serial or a parallel system, until all the elements are completed. This is especially cogent in a parallel system because all elements are being worked on simultaneously. On the other hand, serialistic systems that require a nonzero switching duration from one element to the next must be considered hybrid from the point of this property. Thus, ID is of a bit more technical nature with respect to serial systems than to parallel systems. If the scanning of a memorized list of items in short-term memory can indeed be serial (see especially Chapter 6), it seems likely that the switching time between elements (letters) is minuscule. On the other hand, there are a number of situations where a substantial time to change "focus" would be quite plausible. One such situation is where the task demands an alteration in the modality attended to (see, e.g., LaBerge 1973), for instance, the switching of attention

from the touch sense to the hearing sense. Another class of paradigms has observers listen to different information coming into the two ears (dichotic listening experiments – see, e.g., Moray 1970). In both these situations, of course, the usual hypothesized alternative models to the serial are parallel. The "element" here, with respect to the parallel–serial dichotomy, is a channel (left ear, the tactile sense, etc.) rather than the actual material coming in over the individual channel, although the question might again be raised with regard to the latter, more microscopic level.

Divergent properties: IIA–D

Property IIA

To move on to the aspects of processing that are different for parallel vs. serial processing, consider IIA, on the global state description of the system. It states, in the case of serial systems and models, that at any point in time, at most one element can be in a state of partial completion, whereas the same description in the case of parallel models can have any number of partially completed elements. This is a very natural difference in the two types of processing, since processing is proceeding simultaneously in parallel systems but one at a time in serial systems. Of course, in models based on exponential processing time for individual elements, it may be most reasonable to assume that each element is processed as a gestalt or a single macroscopic component. This kind of processing is all-or-none with regard to element level when the exponential time refers to the complete processing of an element, as we typically do in this book. It then links up with the all-or-none model of perceptual activity proposed by Broadbent (1967) and Townsend (1971a, 1971b, 1978) where a stimulus is presumed to be identified correctly with some probability, but with 1 minus that probability no information at all is acquired about the stimulus. On the other hand, the exponential distribution is not unique in this respect. Only distributions that can reasonably be represented as a convolution of micro-level (i.e., component) processes should be used to represent the processing times of multicomponent elements. The gamma is a distribution of this ilk.

It is crucial to be aware that this property refers to the elements and their components as *input to the system*. If the input elements are degraded so that, for instance, some of their components are missing, then it is the collection of remaining components on which the system operates. This will be important in potential applications discussed below.

When the elements are composed of finer-level components, this property (IIA) entails that serial models will have a grosser state space (sigma or event space in stochastic models) than do parallel models, since the latter can possess nonzero probabilities on states involving several partially completed elements, whereas the serial model cannot.

In principle, property IIA might be made the foundation of empirical attempts to test parallelity vs. seriality. This cannot be done in experiments where perfect processing takes place and only RT is observed, because then we lose the ability to employ the underlying diversity in state spaces; it is obscured when only the finishing times are of interest.

However, consider an experimental paradigm that produces an input to a subsystem that is of high integrity (i.e., the representation of each element carries all the normal information, perhaps in the form of components, in an accurate fashion), but that then abruptly removes the input before all the elements are completed. For example, a whole report experiment with five or fewer letters, high stimulus intensity, and an effective poststimulus visual noise mask may satisfy the appropriate conditions. It is necessary that the stimulus array be present long enough so that if $n=1$, report accuracy is essentially perfect, but it need not be assumed that the noise mask erases the perceptual representation of the stimulus, only that it effectively cessates viable processing efforts. The study reported in Chapter 11 is of this variety, but second guesses were not recorded.

If element identification were truly serial, then there should be at most one letter that is in a partial state of completion at any given point in time. If the elements are unrelated to one another, as in the experiment of Chapter 11, then allowing the observers to engage in second-guessing after their first responses should be very helpful if processing is parallel (and the elements are composed of several components) but of little aid in serial processing. Definite predictions can be made in specific instances. Confidence ratings can also be used to perform similar analyses. This kind of design should be most helpful in ruling out serial models. We presently are aware of no such studies in the literature.

On the other side of the coin, some hybrid models can make predictions similar to those of parallel models. Time-sharing models, which allow the processor to flit from one element to the other without necessarily finishing one before going to the next, make just such predictions as the parallel models. They even have the same sigma field description and can sometimes be made mathematically identical to the parallel models (see Chapter 4). This formal similarity may be the reason that Shevell and Atkinson (1974) treat a class of time-sharing models as if they were parallel models, even though they describe distinct material systems. However, when the notion of capacity is also brought into the discussion, it turns out that situations can be found where one or the other is much more intuitively natural; Chapter 4 can be referred to for a discussion of some of these.

Property IIB

Property IIB indicates that in parallel systems, any element can increase its processing status (degree to which it is completed) at any point in

time during a single trial. A serial system, working on one element at a time, can change the processing state of only one element at a time during a given trial at a certain point in time. One has to be a little careful in giving the mathematical side of this property, because for any Δt, there is *some* finite probability that, even in a serial system, both the element being worked on and the next element could be completed, even though only one is being worked on at a time. Therefore, one cannot just say that the probability that an element that is not currently being processed completes processing during Δt equals 0. However, if one conditions on the current element being in a state of incompletion, throughout Δt, then no subsequent element can be worked on. Clearly this is untrue for parallel models. In particular, there is some nonzero probability that any element can completely finish processing during Δt in a parallel system, but not in a serial system, on a given trial.

Property IIB is in a sense the generator of IIA. However, if each element is made up of only a single component, then there is no way to directly represent the property in terms of probabilities on (potentially observable) completion times. Consider the case of two elements. Unless $p=0$ or 1, either finishing order is possible. Suppose one computes the probability that, say, the element in position b gets completed at or before time t, given that the element in position a is not yet completed at time t *and* a gets finished before b. Then tautologically this probability is 0 in both the parallel and serial cases. On the other hand, if one observes partially completed elements as in property IIA, or consequences thereof as in its discussion, then testable aspects may ensue. Moreover, if p should equal 0 or 1, then this property comes out of the manner in which the element completion probabilities are written in terms of the *actual* processing times of parallel or serial systems. It is indirect, though, and can be satisfied in certain cases by serial as well as parallel models.

There is another consequence of IIB. Recall that the parallel systems are assumed to start with the same processing time distribution on each trial; indeed, given a completely specified history at a certain point in time, the processing time distribution on two such trials must be identical. (This restriction does not hold for compound parallel systems and their models; see Chapter 5.) This aspect of parallel models taken in conjunction with the present property means that the different parts of the probability space of a parallel model must be related in a way that need not be satisfied by a serial model (although it might be). The condition that must be satisfied will be referred to as the *principle of stochastic simultaneity*. A brief mathematical realization of this principle follows.

In the following τ will stand for the actual processing time (equivalent to the total completion time of an element) in the case of a parallel model, and t will stand for the actual (serial) processing time (equivalent to the intercompletion time) of a serial model. Analogously, $\mathbf{T}$ and $\mathbf{T}$ will be the respective random variables corresponding to τ and t.

Now consider the probability function on the joint event that serial position a is processed first at time $\tau_a = t_a = \tau$ and the processing time for the element in serial position b is greater than τ. In a parallel model, this derivation involves integrating the joint density on the two processing times over $\tau_b > \tau$ and setting $\tau_a = \tau$ in that same density. By contrast, in a serial model this probability is simply p times the density on serial position a. First consider the parallel functions

$$\bar{G}(\tau, \tau') = P(\mathbf{T}_a \geq \tau \cap \mathbf{T}_b \geq \tau') = \text{joint survivor function on } \tau_a, \tau_b$$

Next, let

$$g_{ab}(\tau, \tau') = P(\tau + d\tau \geq \mathbf{T}_a \geq \tau \cap \tau' + d\tau' \geq \mathbf{T}_b \geq \tau')$$

and further

$$g_a(\tau, \tau') = P(\tau + d\tau \geq \mathbf{T}_a \geq \tau \cap \mathbf{T}_b \geq \tau') = \int_{\tau'}^{\infty} g_{ab}(\tau, \tau_b)\, d\tau_b$$

$$g_b(\tau, \tau') = P(\mathbf{T}_a \geq \tau \cap \tau' + d\tau' \geq \mathbf{T}_b \geq \tau') = \int_{\tau}^{\infty} g_{ab}(\tau_a, \tau')\, d\tau_a$$

Thus, $g_{ab}(\tau, \tau')$ is the joint density function on the τ_a and τ_b dimensions. It follows that when $\tau' = \tau$, the probability function corresponding to the "event" $\mathbf{T}_a = \tau$ and $\mathbf{T}_b \geq \tau$ is just $g_a(\tau, \tau)$ and that corresponding to $\mathbf{T}_a \geq \tau$ and $\mathbf{T}_b = \tau$ is $g_b(\tau, \tau)$.

We are now in a position to demonstrate that $g_b(\tau, \tau)$ can be directly derived from $g_a(\tau, \tau)$. First, differentiate $g_a(\tau, \tau')$ with respect to τ':

$$\frac{\partial}{\partial \tau'} g_a(\tau_a, \tau') = \frac{\partial}{\partial \tau'} \left[\int_{\tau'}^{\infty} g_{ab}(\tau_a, \tau_b)\, d\tau_b \right]$$

$$= -g_{ab}(\tau_a, \tau')$$

From this we immediately arrive at the desired result by integration:

$$g_b(\tau, \tau) = \int_{\tau}^{\infty} g_{ab}(\tau_a, \tau)\, d\tau_a = \int_{\tau}^{\infty} -\frac{\partial}{\partial \tau} \left[\int_{\tau}^{\infty} g_{ab}(\tau_a, \tau_b)\, d\tau_b \right] d\tau_a$$

Now $g_a(\tau, \tau)$ may similarly be represented as a transformation of $g_b(\tau, \tau')$. The above demonstration is permitted because the joint measure on the parallel probability space involves aspects of both serial positions at all points in space.

Contrarily, once the order is established in a serial system, the density on a has nothing to do with b, and furthermore, the probability that the total completion time on b, τ_b, is greater than τ_a is equal to 1 by definition.

An elementary example that satisfies stochastic simultaneity is found in an independent parallel model where both a and b have gamma distributions with parameters k_a, v_a and k_b, v_b, respectively. (Nonindependent examples can be constructed with a little more effort and tedium.) The germane probability function is

$$g_a(\tau,\tau)=\frac{(v_a\tau)^{k_a-1}v_a\exp(-v_a\tau)}{(k_a-1)!}\sum_{j=k_b+1}^{\infty}\frac{(v_b\tau)^j\exp(-v_b\tau)}{j!}$$

$$=\frac{(v_a\tau)^{k_a-1}v_a\exp(-v_a\tau)}{(k_a-1)!}\int_{\tau}^{\infty}\frac{(v_b\tau_b)^{k_b-1}v_b\exp(-v_b\tau_b)}{(k_b-1)!}\,d\tau_b$$

Carrying out the required partial differentiation and integration yields

$$g_b(\tau,\tau)=\int_{\tau}^{\infty}g_{ab}(\tau_a,\tau)\,d\tau_a=-\int_{\tau}^{\infty}\frac{\partial g_a(\tau_a,\tau')}{\partial\tau'}\bigg|_{\tau'=\tau}d\tau_a$$

$$=\frac{(v_b\tau)^{k_b-1}v_b\exp(-v_b\tau)}{(k_b-1)!}\sum_{j=k_a+1}^{\infty}\frac{(v_a\tau)^j\exp(-v_a\tau)}{j!}$$

Clearly, manipulation of the joint probability space leads from the one expression to the other.

An even simpler serial example that does not, in general, satisfy stochastic simultaneity is our old friend the exponential, with two distinct rate parameters for a and b when either is processed during stage 1 (i.e., $u_{a1}\neq u_{b1}$). Writing the probability function and then transforming with the usual parameter mappings gives

$$P(\tau+d\tau\geqslant\mathbf{T}_a\geqslant\tau\cap\mathbf{T}_b\geqslant\tau)=pu_a\exp(-u_a\tau)$$

$$=\frac{v_{a1}(v_{a1}+v_{b1})\exp[-(v_{a1}+v_{b1})\tau]}{v_{a1}+v_{b1}}$$

$$=v_{a1}\exp[-(v_{a1}+v_{b1})\tau]$$

Now let us observe the alternate probability function when b is first and τ_a is greater than τ. We then transform to a parallel form, but we must allow that $v_a'\neq v_a$ and $v_b'\neq v_b$, where the primed rates are those associated with the transformation when b is first.

$$P(\mathbf{T}_a>\tau\cap\tau+d\tau\geqslant\mathbf{T}_b\geqslant\tau)=(1-p)u_{b1}\exp(-u_{b1}\tau)=v_{b1}'\exp[-(v_{a1}'+v_{b1}')\tau]$$

It is evident in this latter expression that, in general, the two serial probability functions need not be transformable into one another by simple integration and differentiation. Of course, if we constrain $u_{a1}=u_{b1}=u_1=v_{a1}+v_{b1}$ and $v_{a1}=v_{a1}'$, $v_{b1}=v_{b1}'$, then such a transformation is possible by separating the hypothetical contributions from positions a and b according to their association with a and b parameters:

$$pu_{a1}\exp(-u_{a1}\tau)=pu_1\exp(-u_1\tau)=\frac{v_{a1}}{v_{a1}+v_{b1}}(v_{a1}+v_{b1})\exp[-(v_{a1}+v_{b1})\tau]$$

$$=v_{a1}\exp(-v_{a1}\tau_a)\cdot\exp(-v_{b1}\tau_b)\Big|_{\substack{\tau_a=\tau\\ \tau_b=\tau}}$$

and it becomes apparent how the serial space is reconstituted into a parallel

space. In this particular example, the serial ability to depend on future order (with regard to serial positions), discussed in property IID, is employed.

As interesting as the principle of stochastic simultaneity seems theoretically, at present its utility in empirically testing serial and parallel models appears problematic.

Property IIC

Consider next property IIC as applied to serial systems: The time taken to process an element cannot depend on the *identity* of any element that is not completed until after the given element is finished. It contrasts nicely with the nature of parallel activity, where such a dependence can exist. The probability distribution in the two cases can differ because of this. There are a number of ways of illustrating this; we shall indicate one.

Suppose $n=3$ and I_i $(i=1,2,3)$ gives the identity of the elements, where as usual the stimulus display position is just designated by $j=a,b,c$, and $I_i(j)$ then indicates that the element of identity I_i is in position j. Let t_{jk} be the serial actual processing time (and therefore the intercompletion time) for the kth element completed when it is in position j and let f_{jk} be the corresponding density.

Now, consider the situation when only two elements are presented for processing, and suppose that on one type of trial, I_2 is placed by the experimenter in position b [giving $I_2(b)$] and on the other type of trial I_3 is placed in position b [giving $I_3(b)$], while I_1 is in serial position a on both types of trial. Note that display position c is empty and irrelevant. Let us focus on the trials when the processing order follows display position order (for simplicity, and without loss of generality), that is, let $O_1=\langle I_i(a),I_j(b)\rangle$, where $i,j=1,2$ or 3 and $i\neq j$ so that display position a is processed first and display position b is processed second. On these trials the identity indices are $i=1$ and $j=2$ or 3 depending on the trial type. We can now write the serial densities on the first actual processing time in the two cases:

$$f_{a1}(t_{a1}\,|\,O_1,I_1,I_2)\equiv f_{a1}(t_{a1}\,|\,O_1,I_1,I_3)$$

where as usual $\equiv$ means equivalence on all $t_{a1}\in[0,+\infty)$ (except possibly on a set of measure 0). Notice that the order according to serial position is the same, but different elements are processed second in the two cases. According to property IIC the two densities on the first completion time have to be the same in serial models, because a serial *system* has no information concerning the identity of elements processed later.

Now consider the parallel situation. The variable τ will again represent the actual processing time of an element, and therefore its total completion time, given the structure associated with parallel processing. What we want to show is that there are cases where the parallel model *can* yield differences between the two conditioned distributions; it is not necessary that the differences be

present in all distributions. Let A represent the event of I_1 placed in serial position a and I_2 placed in serial position b. Similarly, B also represents the placement of I_1 in serial position a, but I_3 is placed in serial position b.

Now let $g_A(\tau_a, \tau_b)$ be the joint density function on τ_a, τ_b in the first case and $g_B(\tau_a, \tau_b)$ in the second. Also, let $g_{a1}(\tau_{a1} \mid O_1, I_1, I_j)$ be the conditional density on a when it is completed first and I_j $(j=2,3)$ is in serial position b.

It follows from elementary probability theory that

$$g_{a1}(\tau_{a1} \mid O_1, I_1, I_2) = \frac{\int_{\tau_{a1}}^{\infty} g_A(\tau_{a1}, \tau_{b2})\, d\tau_{b2}}{\int_0^{\infty}\int_{\tau_{a1}}^{\infty} g_A(\tau_{a1}, \tau_{b2})\, d\tau_{b2}\, d\tau_{a1}}$$
$$\neq \frac{\int_{\tau_{a1}}^{\infty} g_B(\tau_{a1}, \tau_{b2})\, d\tau_{b2}}{\int_0^{\infty}\int_{\tau_{a1}}^{\infty} g_B(\tau_{a1}, \tau_{b2})\, d\tau_{b2}\, d\tau_{a1}}$$
$$= g_{a1}(\tau_{a1} \mid O_1, I_1, I_3)$$

Where, as usual, the $\neq$ sign indicates that the two expressions have to differ for at least one value of t. [Actually, because we are assuming densities, $\neq$ implies a difference in an infinite number of values (e.g., any interval on $\mathbf{R}^+$), due to the continuity inherent in densities.] The above expressions correspond to the serial conditional densities and prove the assertion if examples satisfying the above inequality can be found.

As in most of the demonstrations we put forth in this chapter, the tried and true exponential distribution can be used for this purpose. Let the identities be $I_1=\alpha$, $I_2=\beta$, $I_3=\gamma$, $u_{ij}(I_k)$ = the serial processing rate with serial position $i=a,b$ and j=stage, and $v_{ij}(I_k)$ = the comparable parallel processing rate; set $p=p(O_1)$. Then in the case of a serial model that obeys the constraint of IIC,

$$f_{a1}(t_{a1} \mid O_1, \alpha, \beta) = u_{a1}(\alpha) \exp[-u_{a1}(\alpha)t]$$
$$= f_{a1}(t_{a1} \mid O_1, \alpha, \gamma) = u_{a1}(\alpha) \exp[-u_{a1}(\alpha)t]$$

In a parallel model, assuming $v_{b1}(\beta) \neq v_{b1}(\gamma)$, that is, that the processing rate varies according to the identity of element β vs. γ, the density with β in position b is

$$g_{a1}(\tau_{a1} \mid O_1, \alpha, \beta)$$
$$= \frac{\int_{\tau_{a1}}^{\infty} v_{a1}(\alpha) \exp[-v_{a1}(\alpha)\tau_{a1}] v_{b1}(\beta) \exp[-v_{b1}(\beta)\tau_{b2}]\, d\tau_{b2}}{\int_0^{\infty}\int_{\tau_{a1}}^{\infty} v_{a1}(\alpha) \exp[-v_{a1}(\alpha)\tau_{a1}] v_{b1}(\beta) \exp[-v_{b1}(\beta)\tau_{b2}]\, d\tau_{b2}\, d\tau_{a1}}$$
$$= \frac{v_{a1}(\alpha) \exp[-v_{a1}(\alpha)\tau_{a1}] \exp[-v_{b1}(\beta)\tau_{a1}]}{v_{a1}(\alpha)/[v_{a1}(\alpha)+v_{b1}(\beta)]}$$
$$= [v_{a1}(\alpha)+v_{b1}(\beta)] \exp[-(v_{a1}(\alpha)+v_{b1}(\beta))\tau_{a1}]$$

Now compare the above expression with the one for element γ in position b:

$$[v_{a1}(\alpha)+v_{b1}(\beta)] \exp[-(v_{a1}(\alpha)+v_{b1}(\beta))\tau_{a1}]$$
$$\stackrel{?}{=} (v_{a1}(\alpha)+v_{b1}(\gamma)) \exp[-(v_{a1}(\alpha)+v_{b1}(\gamma))\tau_{a1}]$$

After some simplification we see that equality will hold only when

$$\tau_{a1} = \frac{\ln\{[v_{a1}(\alpha)+v_{b1}(\beta)]/[v_{a1}(\alpha)+v_{b1}(\gamma)]\}}{[v_{b1}(\beta)-v_{b1}(\gamma)]}$$

That is, equivalence cannot hold for general values of τ_{a1}.

At the risk of stupefying redundancy, note again that the order of processing relative to the serial position was unchanged; only the identity of the element in processing position b (i.e., serial position b in the present case) was altered.

Property IID

The property IID, referring to potential dependence of processing times on a preordained order, has been much discussed in previous chapters. However, it is not always simple to use this as a discriminative principle in the more general models. It works well enough in models with exponential interpletion times and is the basis for the greater generality claimed for serial models in earlier work (e.g., Theorem 1 in Townsend 1972: 174–79). Thus, in exponential models, the serial rate for the first serial position processed can depend on the entire order of processing (again, in terms of the serial position index), whereas the starting set of parallel rates is, of course, the same from trial to trial. In fact, the serial distribution on the minimum completion time will be a probability mixture of exponential distributions (which is not, in general, exponential), but the parallel distribution will be again an exponential distribution with minimum completion time rate given by the sum of all the rates. These remarks clearly apply to later stages of processing up until the last stage, where the parallel rate gains full generality with regard to processing order due to the previously evolved order of completion. An immediate generalization is discovered when we have access to the expression of the probability distributions for the two types of models *and* both types are written as functions of parameters, for then the parallel model's parameters cannot, by fiat, depend on the later order of completion (with respect to serial position), whereas the serial model can.

It would seem natural to further extend these ideas to nonparametrized models by requiring that the part of the parallel probability space corresponding to elements that happen to be completed earlier *not* be able to depend on the later order of completion. Unfortunately, we have not yet worked out a completely general way of expressing this condition mathematically.

One plausible strategy might be to demand that the density on the early completion times be conditionally independent of later events – that is, that such a conditional density be constant across subsequent completion orders. This strategy will not work, however. The problem can be illustrated even within the context of parameterized models, for which an appropriate interpretation of the property does exist, as mentioned above.

As noted, a parallel model cannot make those of its parameters that are applicable in the early stages of processing depend on the later order of com-

pletion. However, the distribution of the first completion time (on, say, serial position a) can, when conditioned on later order, still be different for two different orders (and remember this is all on the serial position; no part of this question concerns the identity of the elements).

Because identity is not important for the present discussion, let us simplify the notation to the old one where the displayed serial positions when $n=3$ are designated as a, b, and c and identity is omitted. However, we will still use τ to refer to the total completion time (=actual processing time in a parallel system) and t for the intercompletion time (=actual processing time in a serial system).

Then the germane serial expressions are

$$f_{a1}(t_{a1}\,|\,O_1)\neq f_{a1}(t_{a1}\,|\,O_2),\quad O_1=\langle a,b,c\rangle,\quad O_2=\langle a,c,b\rangle$$

but in the parallel case it also may be that

$$g_{a1}(\tau_{a1}\,|\,O_1)=\frac{\int_{\tau_{a1}}^{\infty}\int_{\tau'_{b2}}^{\infty}g_{abc}(\tau_{a1},\tau'_{b2},\tau'_{c3})\,d\tau'_{c3}\,d\tau'_{b2}}{\int_0^{\infty}\int_{\tau'_{a1}}^{\infty}\int_{\tau'_{b2}}^{\infty}g_{abc}(\tau'_{a1},\tau'_{b2},\tau'_{c3})\,d\tau'_{c3}\,d\tau'_{b2}\,d\tau'_{a1}}$$

$$\neq g_{a1}(\tau_{a1}\,|\,O_2)$$

$$=\frac{\int_{\tau_{a1}}^{\infty}\int_{\tau'_{c2}}^{\infty}g_{abc}(\tau_{a1},\tau'_{b3},\tau'_{c2})\,d\tau'_{b3}\,d\tau'_{c2}}{\int_0^{\infty}\int_{\tau'_{a1}}^{\infty}\int_{\tau'_{c2}}^{\infty}g_{abc}(\tau'_{a1},\tau'_{b3},\tau'_{c2})\,d\tau'_{b3}\,d\tau'_{c2}\,d\tau'_{a1}}$$

As usual, the fact that, say, τ_b has a subscript attached to it (i.e., 1, 2, or 3) does not mean that the order is predetermined; its purpose is rather to ease the bookkeeping. In a parameterized model, therefore, even though the parameters associated with τ_{a1} do not depend on the later order of processing b and c, the conditional probability density $g(\tau_{a1}\,|\,O_i)$ can depend on O_i in a Bayesian fashion. To paraphrase, we are not claiming that the above nonequivalence ($\neq$) has to hold (in the exponential case, it does not); only that it can hold in general. Thus, a completely general interpretation in terms of the conditional approach is not at present feasible.

Nevertheless, there is a class of models for which a new conditional approach can be legitimately developed. The approach involves breaking down the state-space architecture into the finest-grained description that can in any way affect the processing times of the elements in the various serial positions. At this fine-grain level, the parallel processing must be within-stage-independent at the microstages appropriate to that particular level. Unless one goes down to this critical microstructure, there might be events that, when averaged over in obtaining the pertinent conditional expressions, produce Bayesian dependencies, as in the above example.

For the following development, let *macrostage* refer to the usual interval between completions of two successive elements and *microstage* to the interval between completions of two successive components. Thus, if there were n elements each composed of k components, then exhaustive processing would require nk microstages. Within this constraint, we can show that conditioning on the order and the times of completion of the (micro)*components* will leave

terms that are in no way dependent on the later order of processing of components or elements. That is, the parallel expressions will then be independent of future order.

Let a, b, and c denote the display serial positions of three elements and assume that the elements denoted by a and b are composed of two components each but that the c element is made up of only one component. For simplicity, we will continue to identify elements by a, b, and c here. We are going to suppose for convenience that within an element, the components are processed serially (or by some model-equivalent process).

It will prove convenient to employ our oft-used mode of description in terms of intercompletion times. Of course, these can always be written in terms of the distributions on the total completion times. For example, suppose we are looking at the second intercompletion time, of element b, conditioned on the first macrostage completion time of element a: $g_{b2}(t_{b2}\,|\,t_{a1})$. The value of this intercompletion time density is equivalent for any t_{a1} and any t_{b2} to

$$g_{b2}(t_{b2}\,|\,t_{a1})=g_{b2}(\tau_{b2}-\tau_{a1}\,|\,\tau_{a1})=\frac{g_{ab}(\tau_{a1},\tau_{b2})}{\int_{\tau_{a1}}^{\infty}g_{ab}(\tau_{a1},\tau'_{b2})\,d\tau'_{b2}}$$

where $\tau_{a1}=t_{a1}$ and $\tau_{b2}=t_{a1}+t_{b2}$.

We need some notation for the components, so we will write t_{aij} to refer to the intercompletion time of component i ($=1,2$) belonging to element a in *microstage* j in which it is completed. We shall use, analogously, g_{aij} as the density of *component* i associated with *element* a during *microstage* j; so, for instance,

$$g_{b24}(t_{b24}\,|\,t_{a11},t_{b12},t_{a23})$$

is the density of the second component (e.g., feature) of element (e.g., letter) b, finished during the fourth microstage, after, and conditioned on, the first component intercompletion time of a, the first component intercompletion time of b, and the second (and final) component of a, in that order.

Consider now the times and order $\langle t_{a11},t_{b12},t_{a23},t_{b24},t_{c15}\rangle$ of intercompletion times on the components. Note that element a is completed first but not before one of the two components of b is finished. After a is completed b finishes up, followed by the one (and only) component of element c. Finally let $\bar{G}_{ijk}$=survivor function of component j of element i during microstage k.

Thus accoutered, we are in a position to write the joint density

$$\begin{aligned}
&g_{abc}(t_{a11},t_{b12},t_{a23},t_{b24},t_{c15})\\
&\quad=[g_{a11}(t_{a11})\,\bar{G}_{b11}(t_{a11})\,\bar{G}_{c11}(t_{a11})]\\
&\qquad\times[g_{b12}(t_{b12}\,|\,t_{a11})\,\bar{G}_{a22}(t_{b12}\,|\,t_{a11})\,\bar{G}_{c12}(t_{b12}\,|\,t_{a11})]\\
&\qquad\times[g_{a23}(t_{a23}\,|\,t_{a11},t_{b12})\,\bar{G}_{b23}(t_{a23}\,|\,t_{a11},t_{b12})\,\bar{G}_{c13}(t_{a23}\,|\,t_{a11},t_{b12})]\\
&\qquad\times[g_{b24}(t_{b24}\,|\,t_{a11},t_{b12},t_{a23})\,\bar{G}_{c14}(t_{b24}\,|\,t_{a11},t_{b12},t_{a23})]\\
&\qquad\times[g_{c15}(t_{c15}\,|\,t_{a11},t_{b12},t_{a23},t_{b24})]
\end{aligned}$$

There are obviously five microstages but only three macrostages, the latter corresponding, of course, to the completion of the elements themselves. Now take this joint-time expression and condition on all the intercompletion times and their order, and on the times when they do not get completed (whose probabilities are given by the survivor functions), except for the components of element a (but including the order of finishing of components of a in the above conditionalization).

Let $P(\cdot)$ stand for the joint probability function whose arguments (i.e., variables) refer to the microstage events:

> The first components of b and c finish after the first component of a with $b1$ coming second, followed by $b2$; finally the first (and only) component of c is completed. Further, the b and c components are finished at intercompletion times $t'_{b12}=t_{b12}$, $t'_{b24}=t_{b24}$, $t'_{c15}=t_{c15}$.

In the expression $P(\cdot)$ below, semicolons separate microstages. We then achieve the following density on the intercompletion time of element a when conditioned on the *complete* order and intercompletion times of the b and c components:

$$
\begin{aligned}
&g_{a1}(t_{a11}, t_{a23} \mid \langle a11, b12, a23, b24, c15\rangle; t_{b12}, t_{b24}, t_{c15})\\
&\quad = g_{abc}(t_{a11}, t_{b12}, t_{a23}, t_{b24}, t_{c15})\\
&\qquad + P(\mathbf{T}_{b11} > t_{a11}, \mathbf{T}_{c11} > t_{a11}; \mathbf{T}_{a22} > t_{b12}, \mathbf{T}_{c12} > t_{b12}, \mathbf{T}_{b12} = t_{b12};\\
&\qquad\quad \mathbf{T}_{b23} > t_{a23}, \mathbf{T}_{c13} > t_{a23}; \mathbf{T}_{c14} > t_{b24}, \mathbf{T}_{b24} = t_{b24}; \mathbf{T}_{c15} = t_{c15})\\
&\quad = g_{abc}(t_{a11}, t_{b12}, t_{a23}, t_{b24}, t_{c15})\\
&\qquad + \{[\bar{G}_{b11}(t_{a11})\bar{G}_{c11}(t_{a11})][\bar{G}_{a22}(t_{b12} \mid t_{a11})\bar{G}_{c12}(t_{b12} \mid t_{a11})g_{b12}(t_{b12} \mid t_{a11})]\\
&\qquad\quad \times [\bar{G}_{b23}(t_{a23} \mid t_{a11}, t_{b12})\bar{G}_{c13}(t_{a23} \mid t_{a11}, t_{b12})]\\
&\qquad\quad \times [\bar{G}_{c14}(t_{b24} \mid t_{a11}, t_{b12}, t_{a23})g_{b24}(t_{b24} \mid t_{a11}, t_{b12}, t_{a23})]\\
&\qquad\quad \times [g_{c15}(t_{c15} \mid t_{a11}, t_{b12}, t_{a23}, t_{b24})]\}\\
&\quad = g_{a11}(t_{a11})g_{a23}(t_{a23} \mid t_{a11}, t_{b12})
\end{aligned}
$$

The numerator was already given in its broken-down form above, and it is easily seen that all the terms in the denominator cancel out with an identical term in the numerator, leaving the final expression on the right. Observe that $g_{a11}(t_{a11})g_{a23}(t_{a23} \mid t_{a11}, t_{b12})$ depends only on events up to the time of completion of the second (and final) component of a. Therefore, the overall conditional density on completion of a does not depend on future events, even in a Bayesian fashion. It may seem tautological that this would be so, but if we alter the above conditions at all, in particular the within-microstage independence condition, it will not go through. A special model that fulfills the above condition, yet is not too limited, is one where the g densities are exponential but with the rates at each microstage depending on the previous intercompletion times. Thus, if t_{a11} were the time taken to complete microstage 1 and the first component of element a was the component finished, then we might let the second-stage rate on the first component of (uncompleted) element b be equal to $v_b(t_{a11})$, that is, to be a function of the first-microstage time. Then, in the special linear case that $v_b(t_{a11})=t_{a11}v_b$, the second-stage processing on b would be faster the slower the first stage. This model would satisfy the within-microstage independence condition that is critical to the above derivation.

The principle of correspondent change

We now come to a principle that is held in common with every mathematically specified model or theory in any field. We refer to it as the *principle of correspondent change*. It refers to the regular and predictable manner in which alteration of empirical circumstances should be met with corresponding changes in the model or theoretical structure. One aspect of this is *parameter invariance* (discussed, for example, by Estes 1959) wherein the parameters of a true (or at least more correct) model should not change capriciously; or again, certain parameters should remain constant across certain experimental variations. Thus, it earlier proved possible, within limits, to discover learning parameters (e.g., the rate of learning or conditioning) that were then found to be relatively invariant across certain specified manipulations in the task environment.

In the present context, if, say, the serial rate u were estimated to be a certain value for letters in one task, then it might be expected that the same u should be applicable in other situations in which the mode of letter processing were unaltered. Modeling has rarely reached even this kind of modest goal in memory- and display-scanning research.

In any case, to return to the more general conspectus, the other side of the coin is the regular and orderly change of structure. In the aforementioned learning models, if the items to be learned were made more difficult or the learner rendered (or sampled) more obtuse, then the rate parameter should decrease.

Similarly, consider two serial positions a and b, and suppose we are trying to decide whether a serial or a parallel model better predicts and describes the processing behavior on elements placed in those two positions. Each model will certainly possess structure distinctively associated with the positions. Typically the structure will be in the form of functions of parameters. Now let the experimenter manipulate the complexity of the element placed in, say, position b while leaving the complexity of the element in a the same. Then surely the structure associated with position a should remain invariant in the better model, whereas the structure connected with position b should vary in an appropriate fashion. Of course, dependencies caused, for instance, by lateral interaction or limitations in capacity may affect the rate in position a, but again these should be reflected in regular changes in the "true" model. The wrong model, of course, should exhibit changes of an arbitrary nature in both structures.

To render the discussion a little more concrete, suppose the serial model has processing parameter p, and a and b densities $f_a(t)$ and $f_b(t)$, whereas the parallel model assumes processing is independent with densities $g_a(\tau)$ and $g_b(\tau)$. Imagine that the true model is the serial one. Then, when a solution of the parallel model in terms of the serial model exists, the minimum completion time parallel densities will be of the form (see Chapter 4)

$$g_{a1}(\tau)=\frac{pf_a(\tau)}{p\bar{F}_a(\tau)+(1-p)\bar{F}_b(\tau)}\cdot\exp\left[-\int_0^\tau\frac{pf_a(\tau')}{p\bar{F}_a(\tau')+(1-p)\bar{F}_b(\tau')}\,d\tau'\right]$$

$$g_{b1}(\tau)=\frac{(1-p)f_b(\tau)}{p\bar{F}_a(\tau)+(1-p)\bar{F}_b(\tau)}\cdot\exp\left[-\int_0^\tau\frac{(1-p)f_b(\tau')}{p\bar{F}_a(\tau')+(1-p)\bar{F}_b(\tau')}\,d\tau'\right]$$

As usual, $\bar{F}_i(\tau)$ is 1 minus the distribution function, the so-called survivor function, that is,

$$\bar{F}_i(\tau)=\int_\tau^\infty f_i(\tau')\,d\tau' \quad \text{for } i=a \text{ or } b$$

It is obvious that each of the parallel densities is a function of both the a and b serial distributions. Therefore, the presumed manipulation of the b element should affect only f_b in the fitted serial model, but in the mimicking parallel model both g_a and g_b would be altered.

Some of the results from the section on additivity of reaction times in Chapter 12 may be interpreted in the context of correspondent change. For instance, it was discovered there that in the case of independent parallel processing if experimental variables separately affect the distinct subsystem (as in the minimum-time example just above), then additivity in mean reaction times cannot be predicted. Let X_a and X_b be variables representing the experimental factors affecting subsystems S_a and S_b, respectively. It was shown that the exhaustive processing times of a two-subsystem independent parallel model could be written

$$E(\text{exhaustive processing time}; X_a, X_b)=E[\max(\mathbf{T}_a,\mathbf{T}_b); X_a, X_b]$$
$$=\int_0^\infty [1-G_a(\tau;X_a)G_b(\tau;X_b)]\,d\tau$$

an interactive function of X_a and X_b, whereas a large and natural class of serial models predicted

$$E(\text{exhaustive processing time}; X_a, X_b)=E(\mathbf{T}_1+\mathbf{T}_2; X_a, X_b)$$
$$=p[E(\mathbf{T}_{a1};X_a)+E(\mathbf{T}_{b2};X_b)]$$
$$+(1-p)[E(\mathbf{T}_{b1};X_a)+E(\mathbf{T}_{a2};X_b)]$$
$$=H_a(X_a)+H_b(X_b)$$

that is, an additive function of X_a and X_b as opposed to the interactive parallel function above. This outcome may be viewed as a nonparametric reflection of the principle of correspondent change in the above parallel and serial models.

References

Anderson, J. A. A theory for the recognition of items from short memorized lists. *Psychological Review,* 1973, *80,* 417–438.

Anderson, J. R. *Language, memory, and thought.* Hillsdale, N.J.: Erlbaum, 1976.

Anderson, N. A. Functional measurement and psychological judgment. *Psychological Review,* 1970, *77,* 153–170.

Ashby, F. G. Pattern matching: Self-terminating or exhaustive processing? Unpublished masters thesis, Purdue University, 1976.

Ashby, F. G. Testing the assumptions of exponential, additive reaction time models. *Memory and Cognition,* 1982, *10,* 125–134. (a)

Ashby, F. G. Deriving exact predictions from the cascade model. *Psychological Review,* 1982, *89,* 599–607. (b)

Ashby, F. G. A biased random walk model of two choice reaction times. Submitted to *Journal of Mathematical Psychology,* 1983.

Ashby, F. G., & Townsend, J. T. Decomposing the reaction time distribution: Pure insertion and selective influence revisited. *Journal of Mathematical Psychology,* 1980, *21*(2), 93–123.

Atkinson, R. C., Bower, G. H., & Crothers, E. J. *An introduction to mathematical learning theory.* New York: Wiley, 1965.

Atkinson, R. C., & Estes, W. K. Stimulus sampling theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2). New York: Wiley, 1963.

Atkinson, R. C., Holmgren, J. R., & Juola, J. F. Processing time as influenced by the number of elements in a visual display. *Perception and Psychophysics,* 1969, *6,* 321–326.

Atkinson, R. C., & Juola, J. R. Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 1). San Francisco: Freeman, 1974.

Attneave, F. *Applications of information theory to psychology: A summary of basic concepts, methods, and results.* New York: Holt, Rinehart and Winston, 1959.

Audley, R. J. Some observations on theories of choice reaction time: Tutorial review. In S. Kornblum (Ed.), *Attention and performance* (Vol. 4). New York: Academic Press, 1973.

Audley, R. J., & Pike, A. R. Some alternative stochastic models of choice. *British Journal of Mathematical and Statistical Psychology,* 1965, *18,* 207–225.

Averbach, E., & Sperling, G. Short-term storage of information in vision. In C. Cherry (Ed.), *Symposium on information theory.* London: Butterworth, 1961.

Baddeley, A. D., & Ecob, J. R. Reaction time and short-term memory: Implications of repetition effects for the high-speed exhaustive scan hypothesis. *Quarterly Journal of Experimental Psychology,* 1973, *25,* 229–240.

Ball, K., & Sekuler, R. Models of stimulus uncertainty in motion perception. *Psychological Review,* 1980, *87,* 435–469.

Bamber, D. Reaction times and error rates for "same"–"different" judgments of multidimensional stimuli. *Perception and Psychophysics,* 1969, *6,* 169–174.

Bamber, D. Reaction times and error rates for judging nominal identity of letter strings. *Perception and Psychophysics,* 1972, *12,* 321–326.

Bamber, D. Reaction times in a task analogous to "same"-"different" judgment. *Perception and Psychophysics*, 1975, *18*(5), 321-327.
Barlow, H. B. Retinal noise and absolute threshold. *Journal of the Optical Society of America*, 1956, *46*, 634-639.
Barlow, H. B. Increment thresholds at low intensities considered as signal/noise discriminations. *Journal of Physiology*, 1957, *136*, 469-488.
Bateman, H. The solution of a system of differential equations occurring in the theory of radioactive transformation. *Proceedings of the Cambridge Philosophical Society*, 1910, *15*, 423-527.
Beller, H. K. Parallel and serial stages in matching. *Journal of Experimental Psychology*, 1970, *84*, 213-219.
Berlyne, D. E. Uncertainty and conflict: A point of contact between information theory and behavior-theory concepts. *Psychological Review*, 1957, *64*, 329-339.
Bertelson, P. Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, 1961, *13*, 90-102.
Bertelson, P. Serial choice reaction-time as a function of response versus signal-and-response repetition. *Nature*, 1965, *206*, 217-218.
Bharucha-Reid, A. J. *Elements of the theory of Markov processes and their applications*. New York: McGraw-Hill, 1960.
Bindra, D., Donderi, D. C., & Nishisato, S. Decision latency of "same"-"different" judgments. *Perception and Psychophysics*, 1968, *3*, 121-130.
Bjork, E. L., & Estes, W. K. Detection and placement of redundant signal elements in tachistoscopic displays of letters. *Perception and Psychophysics*, 1971, *9*, 439-442.
Bjork, E. L., & Murray, T. J. On the nature of input channels in visual processing. *Psychological Review*, 1977, *84*, 472-484.
Bloxom, B. Estimating an unobserved component of a serial response time model. *Psychometrika*, 1979, *44*, 473-484.
Booth, T. L. *Sequential machines and automata theory*. New York: Wiley, 1967.
Boring, E. G. *History of experimental psychology*. New York: Appleton-Century-Crofts, 1957.
Borsellino, A., & Poggio, T. Convolution and correlation algebras. *Kybernetik*, 1973, *13*, 113-122.
Bouman, M. A. History and present status of quantum theory in vision. In W. A. Rosenblith (Ed.), *Sensory communication*. Cambridge: MIT Press and Wiley, 1961.
Bower, G. H. Choice-point behavior. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford University Press, 1959.
Broadbent, D. E. Word frequency effect and response bias. *Psychological Review*, 1967, *74*, 1-15.
Brunswik, E. Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 1955, *62*, 193-217.
Burrows, D., & Okada, R. Serial position effects in highspeed memory search. *Perception and Psychophysics*, 1971, *10*, 305-308.
Burrows, D., & Okada, R. Parallel scanning of semantic and formal information. *Journal of Experimental Psychology*, 1973, *97*, 254-257.
Bush, R. R. Estimation and evaluation. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.
Bush, R. R., & Mosteller, F. *Stochastic models for learning*. New York: Wiley, 1955.
Carterette, E. C., Friedman, M. P., & Cosmides, R. Reaction-time distributions in the detection of weak signals in noise. *Journal of the Acoustical Society of America*, 1965, *38*, 531-542.
Chandler, J. P. STEPIT: Finds local minima of a smooth function of several parameters. *Behavioral Science*, 1959, *14*, 81-82.

Chase, W. G., & Calfee, R. C. Modality and similarity effects in short-term recogn tion memory. *Journal of Experimental Psychology*, 1969, *81*, 510-514.
Chase, W. G., & Clark, H. H. Mental operations in the comparison of sentences ar pictures. In L. Gregg (Ed.), *Cognition in learning and memory*. New Yor Wiley, 1972.
Christie, L. S., & Luce, R. D. Decision structure and time relations in simple choi behavior. *Bulletin of Mathematical Biophysics*, 1956, *18*, 89-111.
Churchill, R. V. *Operational mathematics*. New York: McGraw-Hill, 1958.
Clifton, C., & Birenbaum, S. Effects of serial position and delay of probe in a memo scan task. *Journal of Experimental Psychology*, 1970, *86*, 69-76.
Coffman, E. G. (Ed.). *Computer and job-shop scheduling theory*. New York Wiley, 1976.
Cohen, M., & Narens, L. Fundamental unit structures in the theory of measurement *Journal of Mathematical Psychology*, 1980, *20*, 193-232.
Coltheart, M., & Curthoys, I. Short-term recognition memory for pitch: Effect o a priori probability on response times and error rates. *Perception and Psycho physics*, 1968, *4*, 85-89.
Conway, R. W., Maxwell, W. L., & Miller, L. W. *Theory of scheduling*. Reading Mass.: Addison-Wesley, 1967.
Coombs, C. H., Dawes, R. M., & Tversky, A. *Mathematical psychology*. Englewoo Cliffs, N. J.: Prentice-Hall, 1970.
Corballis, M. C., Kirby, J., & Miller, A. Access to elements of a memorized list *Journal of Experimental Psychology*, 1972, *94*, 185-190.
Courant, R. *Differential and integral calculus* (Vol. 2). New York: Interscience, 1936
Cox, D. R. *Renewal theory*. London: Methuen, 1962.
Cox, D. R., & Miller, H. D. *The theory of stochastic processes*. London: Methuen 1965.
Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 671-684.
Creelman, C. D. Human discrimination of auditory duration. *Journal of the Acoustical Society of America*, 1961, *34*, 582-593.
Darley, C. F., Klatzky, R. L., & Atkinson, R. C. Effects of memory load on reaction time. *Journal of Experimental Psychology*, 1972, *96*, 232-234.
Donders, F. C. Die Schnelligkeit Psychischer Processe. *Archiv für Anatomie und Physiologie und Wissenschaftliche Medizin*, 1868, 657-681.
Donders, F. C. On the speed of mental processes. Translated by W. G. Koster in W. G. Koster (Ed.), *Attention and performance* (Vol. 2). Amsterdam: North Holland, 1969.
Edwards, W. Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 1965, *2*, 213-329.
Egeth, H. Parallel versus serial processes in multidimensional stimulus discrimination. *Perception and Psychophysics*, 1966, *1*, 245-252.
Egeth, H., Jonides, J., & Wall, S. Parallel processing of multielement displays *Cognitive Psychology*, 1972, *3*, 674-698.
Ellis, S. H., & Chase, W. G. Parallel processing in item recognition. *Perception and Psychophysics*, 1971, *10*, 379-384.
Eriksen, C. W., & Lappin, J. S. Internal perceptual system noise and redundancy in simultaneous inputs in form identification. *Psychonomic Science*, 1965, *2*, 351-2.
Eriksen, C. W., & Spencer, T. Rate of information processing in visual perception: Some results and methodological considerations. *Journal of Experimental Psychology Monograph*, 1969, *79*, Vol. 2, Part 2.
Estes, W. K. Toward a statistical theory of learning. *Psychological Review*, 1950, *57*, 94-107.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 2). New York: McGraw-Hill, 1959.

Estes, W. K. Interactions of signal and background variables in visual processing. *Perception and Psychophysics*, 1972, *12*, 278–286.

Estes, W. K., Allmeyer, D. H., & Reder, S. M. Serial position functions for letter identification at brief and extended exposure durations. *Perception and Psychophysics*, 1976, *19*, 1–15.

Estes, W. K., & Taylor, H. A. A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences*, 1964, *52*, 446–454.

Estes, W. K., & Taylor, H. A. Visual detection in relation to display size and redundancy of critical elements. *Perception and Psychophysics*, 1966, *1*, 9–16.

Estes, W. K., & Wessel, D. L. Reaction time in relation to display size and correctness of response in forced-choice visual signal detection. *Perception and Psychophysics*, 1966, *1*, 369–373.

Falmagne, J. C. Stochastic models for choice reaction time with applications to experimental results. *Journal of Mathematical Psychology*, 1965, *2*, 77–124.

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

Falmagne, J. C. Random conjoint measurement and loudness summation. *Psychological Review*, 1976, *83*(1), 65–79.

Falmagne, J. C. On a class of probabilistic conjoint measurement models: Some diagnostic properties. *Journal of Mathematical Psychology*, 1979, *19*, 73–89.

Falmagne, J. C., Cohen, S. D., & Dwivedi, A. Two-choice reactions as an ordered memory scanning process. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance* (Vol. 5). London: Academic Press, 1975.

Falmagne, J. C., & Theios, J. On attention and memory in reaction time experiments. *Acta Psychologica*, 1969, *30*, 316–323.

Feller, W. *An introduction to probability theory and its applications.* New York: Wiley, 1957.

Fine, T. *Theories of Probability*, New York: Academic Press, 1973.

Fulkerson, D. R. Expected critical path length in PERT networks. *Operations Research*, 1962, *10*, 808–817.

Gabor, D. Holographic model of temporal recall. *Nature*, 1968, *217*, 584.

Gabor, D. Associative holographic memories. *IBM Journal of Research and Development*, 1969, *13*, 156–159.

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

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

Gardner, G. T. Evidence for independent parallel channels in tachistoscopic perception. *Cognitive Psychology*, 1973, *4*, 130–155.

Gilford, R. M., & Juola, J. F. Familiarity effects on memory and visual search. *Bulletin of the Psychonomic Society*, 1976, *7*, 142–144.

Gilmore, G. C. The interaction among letters in a visual display. *Technical Report 77*, Johns Hopkins University, 1975.

Glezer, V. D., & Nevskaya, K. A. Simultaneous and successive analysis of information in the optic system. Translated from *Doklady Academii USSR*, 1964, *155*(3), 711–714.

Green, D. M. Psychoacoustics and detection theory. *Journal of the Acoustical Society of America*, 1960, *32*, 1189–1203.

Green, D. M., & Luce, R. D. Detection of auditory signals presented at random times. *Perception and Psychophysics*, 1967, *2*, 441–449.

Green, D. M., & Luce, R. D. Detection of auditory signals presented at random times, III. *Perception and Psychophysics*, 1971, *9*, 257–268.

Green, D. M., & Luce, R. D. Speed–accuracy tradeoff in auditory detection. In S. Kornblum (Ed.), *Attention and performance* (Vol. 4). New York: Academic Press, 1972.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* Huntington, N.Y.: Krieger, 1966.

Grossberg, S. Embedding fields: A theory of learning and physiological implications. *Journal of Mathematical Psychology*, 1969, *6*, 209–239.

Grossberg, S. Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, 1978, *17*, 199–219.

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

Hamilton, W. *Lectures on metaphysics and logic* (Vol. 1). Edinburgh: Blackwood, 1859.

Harary, F., Norman, R. Z., & Cartwright, D. *Structural models: An introduction to the theory of directed graphs.* New York: Wiley, 1965.

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

van der Heijden, A. H. C. Some evidence for a limited capacity parallel self-terminating process in simple visual search tasks. *Acta Psychologica*, 1975, *39*, 21–41.

van der Heijden, A. H. C., & Menckenberg, H. W. Some evidence for a self-terminating process in simple visual search tasks. *Acta Psychologica*, 1974, *38*, 169–181.

Hesse, H. *The glass bead game (magister ludi).* New York: Holt, Rinehart and Winston, 1969. (First published in German as *Das Glasperlenspiel.* Zurich: Fretz & Wasmuth, 1943.)

Hick, W. E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 1952, *4*, 11–26.

Hinton, G. E., & Anderson, J. A. (Eds.) *Parallel models of associative memory.* Hillsdale, N.J.: Erlbaum, 1981.

Hohle, R. Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, 1965, *69*, 382–386.

Howell, P., & Stockdale, J. E. Memory and display search in binary classification reaction time. *Perception and Psychophysics*, 1975, *18*, 379–388.

Huesman, L. R., & Woocher, F. D. Probe similarity and recognition of set membership. *Cognitive Psychology*, 1976, *1*, 124–62.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Hunt, C. C., & Kuno, M. Background discharge and evoked response of spinal interneurones. *Journal of Physiology*, 1959, *147*, 364–384.

Huntley, H. E. *Dimensional analysis.* New York: Dover, 1967.

Hylan, J. P. The distribution of attention, I. *Psychological Review*, 1903, *10*, 373–403.

Hylan, J. P. The distribution of attention, II. *Psychological Review*, 1903, *10*, 498–533.

Hyman, R. Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 1943, *45*, 188–196.

Jensen, A. An elucidation of Erlang's statistical works through the theory of stochastic processes. In *The life and works of A. K. Erlang.* Copenhagen: Copenhagen Telephone Company, 1948, pp. 23–100.

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

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

Kantowitz, B. H. Channels and stages in human information processing: A limited review. Unpublished manuscript, 1980.

Kelley, J. E. Critical path planning and scheduling: Mathematical basis. *Operations Research*, 1961, *9*, 296–320.
Kirby, N. Sequential effects in choice reaction time. In A. T. Welford (Ed.), *Reaction Times*. New York: Academic Press, 1980.
Klemmer, E. T. Does recoding from binary to octal improve perception of binary patterns? *Journal of Experimental Psychology*, 1964, *67*, 19–21.
Klir, G. J. *An approach to general systems theory*. New York: Van Nostrand Reinhold, 1969.
Kohfeld, D. L. Effects of the intensity of auditory and visual ready signals on simple reaction time. *Journal of Experimental Psychology*, 1969, *82*, 88–95.
Koopman, B. O. The theory of search, Part I: Kinematic bases. *Operations Research*, 1956, *4*, 324–346. (a)
Koopman, B. O. The theory of search, Part II: Target detection. *Operations Research*, 1956, *4*, 503–531. (b)
Koopman, B. O. The theory of search, Part III: The optimum distribution of searching effort. *Operations Research*, 1957, *5*, 613–626.
Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundations of measurement* (Vol. 1). New York: Academic Press, 1971.
Krantz, D. H., & Tversky, A. Conjoint-measurement analysis of composition rules in psychology. *Psychological Review*, 1971, *78*(2), 151–169.
Krueger, L. E. A theory of perceptual matching. *Psychological Review*, 1978, *85*, 278–304.
Krueger, L. E. A model of unidimensional perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance*. 1979, *5*, 277–288.
LaBerge, D. A recruitment theory of simple behavior. *Psychometrika*, 1962, *27*, 375–396.
LaBerge, D. Identification of two components of the time to switch attention: A test of a serial and a parallel model of attention. In S. Kornblum (Ed.), *Attention and performance* (Vol. 4). New York: Academic Press, 1973.
Laming, D. *Information theory of choice reaction times*. New York: Academic Press, 1968.
Link, S. W. The relative judgment theory of two choice response times. *Journal of Mathematical Psychology*, 1975, *12*, 114–135.
Link, S. W. The relative judgment theory of the psychometric function. In J. Requin (Ed.), *Attention and performance* (Vol. 7). Hillsdale, N.J.: Erlbaum, 1978.
Link, S. W., Ascher, D., & Heath, R. Tests of finite state theories of response times. Unpublished manuscript, McMaster University, 1972.
Link, S. W., & Heath, R. A. A sequential theory of psychological discrimination. *Psychometrika*, 1975, *40*, 77–105.
Lively, B. L. Speed/accuracy trade off and practice as determinants of stage durations in a memory search task. *Journal of Experimental Psychology*, 1972, *96*, 97–103.
Lively, B. L., & Sanford, B. J. The use of category information in a memory search task. *Journal of Experimental Psychology*, 1972, *93*, 379–385.
Loève, M. *Probability theory*. Princeton: Van Nostrand, 1955.
Luce, R. D. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959. (a)
Luce, R. D. On the possible psychophysical laws. *Psychological Review*, 1959, *66*, 81–95. (b)
Luce, R. D. The theory of selective information and some of its behavioral applications. In R. D. Luce (Ed.), *Developments in mathematical psychology*. Glencoe, Ill.: Free Press, 1960.
Luce, R. D. Comments on Rozeboom's criticisms of "On the possible psychophysical laws." *Psychological Review*, 1962, *69*, 548–551.
Luce, R. D. Dimensionally invariant numerical laws correspond to meaningful quali-

Luce, R. D., & Green, D. M. Detection of auditory signals presented at random times II. *Perception and Psychophysics*, 1970, *7*, 1–14.
Luce, R. D., & Green, D. M. A neural timing theory for response times and psychophysics of intensity. *Psychological Review*, 1972, *79*, 14–57.
Luce, R. D., & Tukey, J. W. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1964, *1*(1), 1–27.
Lupker, S. J., & Theios, J. Tests of two classes of models for choice reaction times. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *104*, 137–146.
Lupker, S. J., & Theios, J. Further tests of a two-state model for choice reaction times. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 496–504.
Malcolm, D. G., Roseboom, J. H., Clark, C. E., & Fazar, W. Application of a technique for research and development program evaluation. *Operations Research*, 1959, *7*, 646–669.
Massey, E. J., Jr. The Kolomogorov-Smirnov test of goodness of fit. *Journal of American Statistical Association*, 1951, *46*, 68.
McClelland, J. L. On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 1979, *86*, 287–330.
McGill, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.
McGill, W. J. Introduction to counter theory in psychophysics. Unpublished manuscript, 1964.
McGill, W. J. Neural counting mechanisms and energy detection in audition. *Journal of Mathematical Psychology*, 1967, *4*, 351–376.
McGill, W. J., & Gibbon, J. The general-gamma distribution and reaction times. *Journal of Mathematical Psychology*, 1965, *2*, 1–18.
McGill, W. J., & Goldberg, J. P. Pure-tone intensity discrimination and energy detection. *Journal of the Acoustical Society of America*, 1968, *44*, 576–581.
McGillem, C. D., & Cooper, G. R. *Continuous and discrete signal and system analysis*. New York: Holt, Rinehart and Winston, 1974.
Micko, H. C. A psychological scale for reaction time measurement. *Acta Psychologica*, 1969, *30*, 324–335.
Miller, J. O., & Pachella, R. G. Encoding processing in memory scanning tasks. *Memory and Cognition*, 1976, *4*, 501–506.
Moder, J. J., & Phillips, C. R. *Project management with CPM and PERT*. New York: Van Nostrand Reinhold, 1970.
Moray, N. *Attention: Selective processes in vision and hearing*. New York: Academic Press, 1970.
Morin, R. E., DeRosa, D. V., & Stulz, V. Recognition memory and reaction time. *Acta Psychologica*, 1967, *27*, 298–305.
Murdock, B. B. Convolution and correlation in perception and memory. In L. C. Nilsson (Ed.), *Perspectives on memory research*, Hillsdale, N.J.: Erlbaum, 1979.
Narens, L. A general theory of ratio scalability with remarks about the measurement-theoretic concept of meaningfulness. *Theory and Decision*, 1981, *1*, 1–70.
Narens, L., & Luce, R. D. The algebra of measurement. *Journal of Pure and Applied Algebra*, 1976, *8*, 197–233.
Navon, D. Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 1977, *9*, 353–383.
Navon, D., & Gopher, D. On the economy of the human-processing system. *Psychological Review*, 1979, *86*, 214–255.
Neisser, U. *Cognitive psychology*. Englewood Cliffs, N.J.: Prentice-Hall, 1967.
Nickerson, R. S. Response time with memory-dependent decision task. *Journal of*

Nickerson, R. S. Binary-classification reaction time: A review of some studies of human information-processing capabilities. *Psychonomic Monograph Supplements*, 1972, *4*, 275–318.

Noreen, D. L., & Burns, D. Probability effects and stimulus differences in reaction time. Paper presented at the Ninth Annual Mathematical Psychology Meetings, New York University, 1976.

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

Norman, M. F. Lectures on systems theory. *Journal of Mathematical Psychology*, 1981, *23*, 1–89.

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

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

Padulo, L., & Arbib, M. *System theory*. Washington, D.C.: Hemisphere, 1974.

Palm, C. The distribution of repairmen in servicing automatic machines. *Industritidningen Norden*, 1947, *75*, 75–80, 90–94, 119–123. (In Swedish.)

Papoulis, A. *Probability, random variables, and stochastic processes*. New York: McGraw-Hill, 1965.

Parzen, E. *Modern probability theory and its applications*. New York: Wiley, 1960.

Parzen, E. *Stochastic processes*. San Francisco: Holden-Day, 1962.

Patil, G. P., & Joshi, S. W. *A dictionary and bibliography of discrete distributions*. Edinburgh: Oliver and Boyd, 1968.

Pike, A. R. Stochastic models of choice behavior: Response probabilities and latencies of finite Markov chain systems. *British Journal of Mathematical and Statistical Psychology*, 1966, *19*, 15–32.

Pike, A. R. Latency and relative frequency of response in psychophysical discrimination. *British Journal of Mathematical and Statistical Psychology*, 1968, *21*, 161–182.

Pike, A. R. The latencies of correct and incorrect responses in discrimination tasks: Their interpretation in terms of a model based on simple counting. *Perception and Psychophysics*, 1971, *9*, 455–460.

Pike, A. R. Response latency models for signal detection. *Psychological Review*, 1973, *80*, 53–68.

Pollack, I., & Johnson, L. B. Memory span with efficient coding procedures. *American Journal of Psychology*, 1965, *78*, 609–614.

Posner, M. I., & Mitchell, R. F. Chronometric analysis of classification. *Psychological Review*, 1967, *74*, 392–409.

Pribram, K. H. Some dimensions of remembering: Steps toward a neuro-psychological model of memory. In J. Guito (Ed.), *Macromolecules and behavior*. New York: Appleton-Century-Crofts, 1966.

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

Ratcliff, R. Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 1979, *86*, 446–461.

Ratcliff, R., & Murdock, B. B., Jr. Retrieval processes in recognition memory. *Psychological Review*, 1976, *83*, 190–214.

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

Restle, R., & Greeno, J. G. *Introduction to mathematical psychology*. Reading, Mass.: Addison-Wesley, 1970.

Roberts, F. S. *Discrete mathematical models: With applications to social, biological, and environmental problems*. Englewood Cliffs, N.J.: Prentice-Hall, 1976.

Roberts, F. S. *Measurement theory with applications to decisionmaking, utility, and the social sciences*. Reading, Mass.: Addison-Wesley, 1979.

Ross, B. H., & Anderson, J. R. A test of parallel versus serial processing applied to memory retrieval. *Journal of Mathematical Psychology*, 1981, *24*(3), 183–223.

Rossmeissl, P. G., Theios, J., & Krunnfusz, D. Parallel processing models for joint visual and memory scanning. Paper presented at the Twelfth Annual Mathematical Psychology Meetings, Brown University, August, 1979.

Rozeboom, W. W. The untenability of Luce's principle. *Psychological Review*, 1962, *69*, 542–547.

Rumelhart, D. E. A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 1970, *7*, 191–218.

Saaty, T. L., & Bram, J. *Nonlinear mathematics*. New York: McGraw-Hill, 1964.

Sakitt, B. Iconic memory. *Psychological Review*, 1976, *83*, 257–276.

Scheerer, E. Order of report and order of scanning in tachistoscopic recognition. *Canadian Journal of Psychology*, 1972, *26*, 382–390.

Scheerer, E. A further test of the scanning hypothesis in tachistoscopic recognition. *Canadian Journal of Psychology*, 1973, *27*, 95–102.

Scheerer, E. Task requirement and hemifield asymmetry in tachistoscopic partial report performance. *Acta Psychologica*, 1974, *38*, 131–147.

Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 1977, *84*, 1–66.

Schönemann, P. H., & Steiger, J. H. Regression and component analysis. *British Journal of Mathematical and Statistical Psychology*, 1976, *29*, 175–189.

Schweickert, R. A critical path generalization of the additive factor method: Analysis of a stroop task. *Journal of Mathematical Psychology*, 1978, *18*, 105–139.

Sedov, L. I. *Similarity and dimensional methods in mechanics*. Moscow, 1956. English translation by M. Holt & M. Friedman. New York: Academic Press, 1959.

Shannon, C. E., & Weaver, W. *The mathematical theory of communication*. Urbana, Ill.: University of Illinois Press, 1949.

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

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

Shaw, P., & LaBerge, D. Manipulation of visual processing by varying the rewards associated with display locations. *Perception and Psychophysics*, 1971, *10*, 348–352.

Sheridan, T. B., & Ferrel, W. R. *Man-machine systems*. Cambridge: MIT Press, 1974.

Shevell, S. K., & Atkinson, R. A. Theoretical comparison of list scanning models. *Journal of Mathematical Psychology*, 1974, *11*, 76–106.

Shiffrin, R. M., & Gardner, G. T. Visual processing capacity and attentional control. *Journal of Experimental Psychology*, 1972, *93*, 72–82.

Shiffrin, R. M., & Geisler, W. S. Visual recognition in a theory of information processing. In R. L. Solso (Ed.), *Contemporary issues in cognitive psychology: The Loyla symposium*. Washington, D.C.: Winston, 1973.

Shulman, H. G., & McConkie, A. S-R compatibility, response discriminability, and response codes in choice reactions. *Journal of Experimental Psychology*, 1973, *98*, 375–378.

Shulman, G. L., Remington, R. W., & McLean, J. P. Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 522–526.

Siebert, W. M. Some implications of the stochastic behavior of primary auditory neurons. *Kybernetik*, 1965, *2*, 206–215.

Siebert, W. M. Stimulus transformations in the peripheral auditory system. In P. A. Kohlers and M. Eden (Eds.), *Recognizing patterns*. Cambridge: MIT Press, 1968.

Siebert, W. M. Frequency discrimination in the auditory system: Place or periodicity mechanisms? *Proceedings of the IEEE,* 1970, *58,* 723–730.

Sigal, C. E., Pritsker, A. A. B., & Solberg, J. J. The stochastic shortest route problem. *Operations Research,* 1980, *28,* 1122–1129.

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

Snodgrass, J. G., Luce, R. D., & Galanter, E. Some experiments on simple and choice reaction time. *Journal of Experimental Psychology,* 1967, *75,* 1–17.

Snodgrass, J. G., & Townsend, J. T. Comparing parallel and serial models: Theory and implementation. *Journal of Experimental Psychology: Human Perception and Performance,* 1980, *6*(2), 330–354.

Sperling, G. The information available in brief visual presentation. *Psychological Monographs,* 1960, *74*(11) (whole no. 498).

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

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

Sperling, G. Early experience and visual information processing in perceptual and reading disorders. In Proceedings of a Conference in Association with the Committee on Brain Sciences, Division of Medical Science, National Research Council, 1970.

Starke, P. H. *Abstract automata.* Amsterdam: North Holland, 1972.

Sternberg, S. Estimating the distribution of additive reaction time components. Paper presented at the annual meeting of the Psychometric Society, 1964.

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

Sternberg, S. Scanning a persisting visual image versus a memorized list. Paper presented at the Annual Meeting of the Eastern Psychological Association, 1967.

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance* (Vol. 2). Amsterdam: North Holland Press, 1969. (a)

Sternberg, S. Memory-scanning: Mental processes revealed by reaction time experiments. *American Scientist,* 1969, *57*(4), 421–457. (b)

Sternberg, S. Evidence against self-terminating memory search from properties of RT distributions. Paper presented at a meeting of the Psychonomic Society, St. Louis, November, 1973.

Sternberg, S. Memory-scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology,* 1975, *27,* 1–32.

Sternberg, S., & Knoll, R. L. The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance* (Vol. 4). New York: Academic Press, 1973.

Stevens, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *Journal of Experimental Psychology,* 1959, *57,* 201–209.

Stevens, S. S. Matching functions between loudness and ten other continua. *Perception and Psychophysics,* 1966, *1,* 5–8.

Stone, L. D. *Theory of optimal search.* N.Y.: Academic Press, 1975.

Stone, M. Model for choice-reaction time. *Psychometrika,* 1960, *25,* 251–260.

Suppes, P., & Zinnes, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.

Swanson, J. M., & Briggs, G. E. Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology,* 1969, *81,* 223–239.

Swensson, R. G., & Edwards, W. Response strategies in a two-choice reaction task with a continuous cost for time. *Journal of Experimental Psychology,* 1971, *88,* 67–81.

Swensson, R. G., & Green, D. M. On the relations between random walk models for two-choice response times. *Journal of Mathematical Psychology,* 1977, *15,* 282–291.

Tanner, W. P., & Swets, J. A. A decision-making theory of visual detection. *Psychological Review,* 1954, *61,* 401–409.

Tapia, R. A., & Thompson, J. R. *Nonparametric probability density estimation.* Baltimore: Johns Hopkins University Press, 1978.

Taylor, D. A. Questioning the assumption of selective influence in the analysis of search and matching tasks. Paper presented at the Sixth Annual Mathematical Psychology Meetings, Montreal, August 1973.

Taylor, D. A classification scheme for models of multi-element processing. Paper presented at the Eighth Annual Mathematical Psychology Meetings, Purdue University, August 1975.

Taylor, D. A. Holistic and analytic processes in the comparison of letters. *Perception and Psychophysics,* 1976, *20,* 187–190. (a)

Taylor, D. A. Effect of identity in the multiletter matching task. *Journal of Experimental Psychology: Human Perception and Performance,* 1976, *2,* 417–428. (b)

Taylor, D. A. Stage analysis of reaction time. *Psychological Bulletin,* 1976, *83,* 161–191. (c)

Taylor, D. A., Townsend, J. T., & Sudevan, P. Analysis of intercompletion times in multi-element processing. Paper presented at the annual meeting of the Psychonomic Society, San Antonio, November 1978.

Theios, J. Reaction time measurements in the study of memory processes: Theory and data. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7). New York: Academic Press, 1973.

Theios, J., & Smith, P. G. Can a two-state model account for two-choice reaction-time data? *Psychological Review,* 1972, *79,* 172–177.

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

Thomas, E. A. C. Sufficient conditions for monotone hazard rate: An application to latency-probability curves. *Journal of Mathematical Psychology,* 1971, *8,* 303–332.

Thomas, E. A. C. The selectivity of preparation. *Psychological Review,* 1974, *81,* 442–464.

Tolman, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century-Crofts, 1932.

Townsend, J. T. Mock parallel and serial models and experimental detection of these. *Purdue Centennial Symposium on Information Processing,* Purdue Press, Purdue University, April 1969.

Townsend, J. T. Theoretical analysis of an alphabetic confusion matrix. *Perception and Psychophysics,* 1971, *9,* 40–50. (a)

Townsend, J. T. Alphabetic confusion: A test of models for individuals. *Perception and Psychophysics,* 1971, *10,* 449–454. (b)

Townsend, J. T. Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology,* 1972, *25,* 168–199.

Townsend, J. T. Mathematical aspects of finite processing systems. Talk given at Mathematical Psychology Seminar, New York University, February 1973. (a)

Townsend, J. T. Chairman and contributor, Symposium on Models for Search in Short-term Memory, Sixth Annual Mathematical Psychology Meetings, with W. K. Estes & George Sperling (discussants), & R. A. Kinchla, J. Anderson, & P. Cavanaugh (contributors), Montreal, August 1973. (b)

Townsend, J. T. Independent parallel models and analyses of reaction time distributions. Paper presented at Seventh Annual Mathematical Psychology Meetings, Ann Arbor, August 1974. (a)

Townsend, J. T. Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Hillsdale, N.J.: Erlbaum, 1974, pp. 133–168. (b)

Townsend, J. T. A stochastic theory of matching processes. *Journal of Mathematical Psychology*, 1976, *14*, 1–52. (a)

Townsend, J. T. Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology*, 1976, *14*, 219–238. (b)

Townsend, J. T. A clarification of some current multiplicative confusion models. *Journal of Mathematical Psychology*, 1978, *18*(1), 25–38.

Townsend, J. T. Some characteristics of visual whole report behavior. *Acta Psychologica*, 1981, *47*(2), 149–173.

Townsend, J. T., & Ashby, F. G. Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, N.J.: Erlbaum, 1978, pp. 200–239.

Townsend, J. T., & Fial, R. Spatio-temporal characteristics of multisymbol perception. Paper presented at Mathematical Psychology Meetings, Stanford University, August 1968.

Townsend, J. T., & Romei, S. Characteristics of reaction time distributions. Invited presentation at the Formal Theories of Information Processing Conference, Stanford University, June 1974.

Townsend, J. T., & Roos, R. N. Search reaction time for single targets in multiletter stimuli with brief visual displays. *Memory and Cognition*, 1973, *1*, 319–332.

Townsend, J. T., & Snodgrass, J. G. A serial vs parallel testing paradigm when "same" and "different" comparison rates differ. Paper presented at Psychonomic Society Meetings, Boston, November 1974.

Townsend, J. T., & Snodgrass, J. G. Masking effects of similar and dissimilar letters: A test of the interactive channels model. *Perceptual and Motor Skills*, 1977, *44*, 1212–1228.

Townsend, J. T., Taylor, S. G., & Brown, D. R. Lateral masking for letters with unlimited viewing time. *Perception and Psychophysics*, 1971, *10*, 375–378.

Travers, J. R. The effects of forced serial processing on identification of words and random letter strings. *Cognitive Psychology*, 1973, *5*, 109–37.

Tversky, A., & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 1973, *5*, 207–232.

Tversky, B. Pictorial and verbal encoding in a short-term memory task. *Perception and Psychophysics*, 1969, *6*, 225–233.

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

Vorberg, D. On the equivalence of parallel and serial models of information processing. Paper presented at the Tenth Annual Mathematical Psychology Meetings held at University of California at Los Angeles, 1977.

Vorberg, D. Reaction time distributions predicted by serial self-terminating models of memory search. In S. Grossberg (Ed.), *Mathematical Psychology and Psychophysiology*, Proceedings of the Symposium in Applied Mathematics of the American Mathematical Society and the Society for Industrial and Applied Mathematics. New York: American Mathematical Society, 1981.

Wald, A. *Sequential analysis*. New York: Wiley, 1947.

Wald, A., & Wolfowitz, J. Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 1948, *19*, 326–339.

Walsh, J. E. *Handbook of Nonparametric Statistics*. II. New York: Van Nostrand, 1965.

Wiest, J. D., & Levy, F. K. *A management guide to PERT/CPM* (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Wingfield, A., & Bolt, R. A. Memory search for multiple targets. *Journal of Experimental Psychology*, 1970, *85*, 45–50.

Wolford, G., & Hollingsworth, S. Evidence that short-term memory is not the limiting factor in the tachistoscopic full-report procedure. *Memory and Cognition*, 1974, *2*, 796–800.

Wolford, G., Wessel, D., & Estes, W. K. Further evidence concerning scanning and sampling assumptions of visual detection models. *Perception and Psychophysics*, 1968, *3*, 439–444.

Wundt, W. *Lectures on human and animal psychology*. London: Swan Sonneschein, 1894.

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

Yellott, J. I. Correction for fast guessing and the speed–accuracy trade-off in choice reaction time. *Journal of Mathematical Psychology*, 1971, *8*, 159–199.

Author index

Allmeyer, D. H., 334, 348
Anderson, J. A., 133, 403
Anderson, J. R., 61, 415, 457
Anderson, N. A., 357
Arbib, M., 401
Ascher, D., 266, 268
Ashby, F. G., 36, 40–1, 51, 76–7, 92, 133, 140, 153, 166, 179–80, 185–7, 189, 190, 202, 204, 208, 219, 248, 253, 280, 284, 311, 365, 371, 387, 406, 408, 409, 422
Atkinson, R. A., 63–5, 471
Atkinson, R. C., xii, 2, 9, 14, 50, 102, 116, 119, 122, 131, 138, 180, 186, 241, 244, 259, 261
Attneave, F., 6
Audley, R. J., 257, 259, 275
Averbach, E., 318

Baddeley, A. D., 124, 133
Ball, K., 269–70
Bamber, D., 153–5, 170, 173, 417, 422
Barlow, H. B., 257
Bateman, H., 39
Bates, M., 64
Beller, K., 156
Berlyne, D. E., 119
Bertelson, P., 134
Bharucha-Reid, A. J., 39
Bindra, D., 155
Birenbaum, S., 122
Bjork, E. L., 124, 168, 169, 346
Bloxom, B., 153
Bobrow, D. G., 76
Bolt, R. A., 151
Booth, T. L., 401
Boring, E. G., 4
Borsellino, A., 403
Bouman, M. A., 257
Bower, G. H., xiv, 2, 62
Bram, J., 357
Briggs, G. E., 255
Broadbent, D. E., 261, 470
Brown, D. R., 334, 348
Brunswik, E., 102
Burns, D., 266
Burrows, D., 119
Bush, R. R., 2, 6, 44

Calfee, R. C., 120
Carterette, E. C., 259, 263, 308, 314
Cartwright, D., 16
Chandler, J. P., 430, 440
Chase, W. G., 120, 131, 417
Christie, L. S., 16, 36, 40
Churchill, R. V., 35
Clark, C. E., 16
Clark, H. H., 417
Clifton, C., 122
Coffman, E. G., 16
Cohen, M., 397
Cohen, S. D., 264
Coltheart, M., 437
Conway, R. W., 16
Coombs, C. H., xii
Cooper, G. R., 401
Corballis, M. C., 133
Cosmides, R., 259, 263, 308, 314
Courant, R., 102
Cox, D. R., 27, 32–3, 62, 297, 298
Craik, F. I. M., 153
Creelman, C. D., 257
Crothers, E. J., xii, 2
Curthoys, I., 437

Darley, C. F., 131
Dawes, R. M., xii
DeRosa, D. V., 122
Donderi, D. C., 155, 219
Donders, F. C., 5, 145, 219, 358, 363
Dwivedi, A., 264

Ecob, J. R., 124, 133
Edwards, W., 259, 293
Egeth, H., 125, 153
Ellis, S. H., 131
Eriksen, C. W., 84, 320, 346
Estes, W. K., 6, 102, 116, 124, 168–9, 258–9, 317–8, 319, 320, 334, 345–6, 348, 371, 445, 481

Falmagne, J. C., 134, 261, 264, 266, 271, 390
Fazar, W., 16
Feller, W., 38–9, 44–5
Ferrel, W. R., 139
Fial, R., 329, 350
Fine, T., 450
Friedman, M. P., 259, 263, 308, 314
Fulkerson, D. R., 16

Gabor, D., 403

Galambos, J., 92
Galanter, E., 51, 133, 259, 263, 308, 314
Gardner, G. T., 118, 125, 320, 346–7, 355
Garner, W. R., 6
Geisler, W. S., 346
Gibbon, J., 36, 40, 51, 132, 209
Gilford, R. M., 117
Gilmore, G. C., 346
Glezer, V. D., 324
Goldberg, J. P., 257
Gopher, D., 76, 131
Green, D. M., 6, 36, 38, 51, 139, 257, 294, 309
Greeno, J. G., xii, 2
Grossberg, S., 133, 361
Gumbel, E. J., 27, 92, 129

Hamilton, W., 445
Harary, F., 16
Harris, J. R., 64
Haviland, S. E., 120, 134–6
Heath, R., xiv, 266, 268, 293–5, 298, 301, 307, 316
van der Heijden, A. H. C., 124–5, 168–9, 180, 349
Hesse, H., xi
Hick, W. E., 119
Hinton, G. E., 403
Hohle, R., 51
Hollingsworth, S., 346–7
Holmgren, J. R., 9, 50, 116, 119, 122, 131, 138, 180, 186, 241, 244
Howell, P., 152, 155, 156
Huesman, L. R., 417
Hull, C. L., 371
Hunt, C. C., 23
Huntley, H. E., 397
Hylan, J. P., 125, 193
Hyman, R., 134

Jensen, A., 38
Johnson, L. B., 348
Jonides, J., 125
Joshi, S. W., 174, 199
Juola, J. F., 9, 50, 116–17, 119, 122, 131, 138, 180, 186, 241, 244, 261

Kahneman, D. K., 76, 156
Kantowitz, B. H., 76, 131
Kelley, J. E., 16
Kirby, J., 133
Kirby, N., 134
Klatzky, R. L., 131
Klemmer, E. T., 348
Klir, G. J., 401
Knoll, R. L., 5
Kohfeld, D. L., 257
Koopman, B. O., 139
Krantz, D. H., xii, 357, 389–91, 395, 397–400
Krueger, L. E., 155, 170, 416–7, 430, 437, 445
Krunnfusz, D., 152, 201
Kuno, M., 23

LaBerge, D., 109, 111, 257, 275, 469
Laming, D., xiv, 293, 316
Lappin, J. S., 84
Levy, F. K., 16
Link, S. W., xiv, 266, 268, 293, 294, 295, 298, 301, 307, 311, 314–6
Lively, B. L., 128
Lockhart, R. S., 153
Loeve, M., 450
Luce, R. D., xii, 6, 16, 36, 38, 40, 51, 133, 184, 257, 259, 263, 308, 314, 357, 389–91, 394–5, 397–400
Lupker, S. J., 244, 266, 271

Malcolm, D. G., 16
Massey, E. J., Jr., 156
Maxwell, W. L., 16
McClelland, J. L., 371, 408–9
McConkie, A., 119
McGill, W. J., xiii, 33, 36, 39, 40, 44, 51, 132, 209, 257
McGillem, C. D., 401
McLean, J. P., 355
Menckenberg, H. W., 124–5, 169, 180
Micko, H. C., 387, 390
Miller, A., 133
Miller, H. D., 32–3, 62, 297–8
Miller, J. O., 118
Miller, L. W., 16
Mitchell, R. F., 153
Modor, J. J., 16
Moray, N., 470
Morin, R. E., 122
Mosteller, F., 6, 44
Moy, M. C., 120, 134–6
Murdock, B. B., 51, 133, 403
Murray, T. J., 346

Narens, L., 397
Navon, D., 76, 131, 445
Neisser, U., 117
Nevskaya, K. A., 324
Nickerson, R. S., 133, 146, 151, 154, 165
Nishisato, S., 155
Noreen, D. L., 266
Norman, D. A., 76
Norman, M. F., 401
Norman, R. Z., 16

Okada, R., 119
Ollman, R., 256, 259

Pachella, R. G., 118, 255, 293, 371
Padullo, L., 401
Palm, C., 38

Papoulis, A., 78
Parzen, E., 24, 33, 77–9, 194
Patil, G. P., 174, 199
Phillips, C. R., 16
Pike, A. R., 257, 259, 275, 279
Poggio, T., 403
Pollack, I., 348
Posner, M. I., 153
Pribram, K. H., 403
Pritsker, A. A. B., 16

Ratcliff, R., 51, 95, 133, 153, 193
Reder, S. M., 334, 348
Remington, R. W., 355
Restle, F., 36, 40, 132
Restle, R., xii, 2
Roberts, F. S., 16, 389
Romei, S., 457
Roos, R. N., 9, 12, 115, 117, 120, 122, 136, 148, 167, 186, 187, 190, 207, 244, 246–8, 318, 417, 464
Roseboom, J. H., 16
Ross, B. H., 61, 415
Rossmeissl, P. G., 152, 201
Rozeboom, W. W., 397, 399
Rumelhart, D. E., 62, 64–5, 345, 349–50, 351

Saaty, T. L., 357
Sakitt, B., 117
Sanford, B. J., 128
Scheerer, E., 318, 347
Schneider, W., 124, 193, 201
Schönemann, P. H., xvi
Schweickert, R., 15–18, 371
Sedov, L. I., 397
Sekuler, R., 269–70
Shannon, C. E., 76
Shaw, M. L., 64, 139–41, 143–5
Shaw, P., 109, 111, 139
Sheridan, T. B., 139
Shevell, S. K., 63–5, 471
Shiffrin, R. M., 118, 124, 193, 201, 320, 346–7, 355
Shulman, G. L., 355
Shulman, H. G., 119
Siebert, W. M., 257
Sigal, C. E., 16
Smith, E. E., 118
Smith, P. G., 120, 134–6
Snodgrass, J. G., 51, 133, 162, 258–63, 278, 308–9, 314, 346, 415, 430, 437–8, 464
Solberg, J. J., 16
Spencer, T., 320, 346
Sperling, G., 62, 117, 317–19, 321, 329, 331, 334, 345
Starke, P. H., 401
Sternberg, S., 5, 15, 50, 115, 118–22, 128, 130–2, 138–9, 151–2, 168, 190, 198, 201, 207, 218, 220–2, 229, 241–4, 268, 359–61, 363–4, 387, 390
Stevens, S. S., 400
Stockdale, J. E., 152, 155–6
Stone, L. D., 140
Stone, M., 293, 316
Stulz, V., 122
Sudevan, P., 157–8
Suppes, P., xii, 357, 389–91, 395, 397–400
Swanson, J. M., 255
Swensson, R. G., 259, 294, 309
Swets, J. A., 6, 139

Tanner, W. P., 6
Tapia, R. A., 153
Taylor, D. A., 65, 146, 153–8, 165, 173, 370–1, 417
Taylor, H. A., 116, 168, 258, 317–19, 320, 334, 346, 445
Taylor, S. G., 334, 348
Theios, J., 120, 134–6, 152, 201, 244, 266, 271
Thomas, E. A. C., 5, 95
Thompson, J. R., 153
Tolman, E. C., 371
Townsend, J. T., 9, 12, 14, 36, 40–1, 48, 51, 58, 60–2, 63–4, 76–7, 92, 101, 103, 115, 117, 120, 122, 131, 133, 136, 138, 140, 146, 148, 153, 155, 157–8, 162, 167, 174, 186–7, 190, 202, 204, 207–8, 219, 229, 242, 244, 246–8, 250, 253, 256, 258–62, 268, 280, 308–9, 317–21, 329, 334, 345–6, 348–9, 350, 365, 371, 384–5, 387, 406, 414–15, 417–18, 422, 427, 430–1, 437, 457, 464–5, 470, 477
Traupman, J., 120, 134–6
Travers, J. R., 329
Tukey, J. W., 390
Tversky, A., xii, 156, 357, 389–91, 395, 397–400
Tversky, B., 155

Vickers, D., 256
Vorberg, D., 48, 201, 218

Wald, A., 293, 295, 299, 310
Wall, S., 125
Walsh, J. E., 242
Weaver, W., 76
Wessel, D. L., 259, 348
Wiest, J. D., 16
Wingfield, A., 151
Wolford, G., 346–8
Wolfowitz, J., 310
Woocher, F. D., 417
Wundt, W., xi, 5, 358

Yellott, J. I., 256, 259, 265–6

Zinnes, J. L., 389

Subject index

accessibility ordering assumption, 218–32, 268
accuracy, 109–14, 203–5, 255–316, 317–55
actual processing time, 10, 12, 49
additive factor method, 359–87, 395–6, 407, 411
additivity of reaction time, 356–413, 417
  linear system in, 407–8
  mean, 363, 417; measurement theory and, 390–6; parallel systems in, 371–83, 384–7; probabilistic automata, 410–11; serial systems in, 363–71, 383–7
all-or-none model, 256, 261
automata, probabilistic, 409–12

base time, 120, 269
Bayes's theorem, 313
Bernoulli trials, 43
binary mixture models, *see* varied-state models
binomial distribution, 45–6
bivariate distribution, 27–9
bounded performance model, 350–5

capacity, 13–15
  energy and power, 77–80
  exhaustive level, 208–12
  extremely limited, 176–7
  first intercompletion time level, 248–54
  fixed, 85–9, 176–7
  individual element level, 215–18
  limited, 77, 89–91, 176, 216
  reallocation, 48–9, 88–9, 138–45
  self-terminating level, 212–15
  serial models, 80–3
  standard unlimited capacity parallel model, 217
  supercapacity, 76–7, 176, 216–17
  unitary vs. multiple sources, 131–2
  unlimited, 76, 83–5, 176, 216–17
  whole report, 345–50
Cauchy distribution, 33–4
characteristic function, 34
chi-square statistic, 2
choice rule, 184–5
comparison process, 118
completion order, 53, 64
completion time, 49
compound processing models, 99–114
conjoint measurement, 390–6
convolution, 30–2, 269, 403–6
counting models, 256–7, 272–90, 292
  counting rate vs. criterion effects, 285–8
  processing time density function, conditional, 276–7
  response probabilities, unconditional, 274–5
counting process, 77–80, 284–5

density function, 24–6, 31, 35–6, 42, 45
  reaction time density, 132–3, 153, 266–71
deterministic models, 8–22, 23, 64
  latent network theory, 15–22
dimensional analysis, 396–401
discrete vs. continuous distribution, 43–4
display configuration, 125–6
distribution function, 24–6, 34–6, 45
distribution ordering, 170–1, 206–54, 281–3
dominance relations, 208, 248, 280–5
Donderian systems, 5–6, 358, 408, 410

early target paradigms, *see* visual search
Erlang distribution, 40–1, 209
exhaustive processing, 12–13, 65–8, 164–205, 208–12
expectation, 29–30, 45
exponential distribution, 36–8, 51–2
  convolution of, 209–10
  counting models, 273–90
  memory and display search, criticism of, 132–3
  parallel serial equivalence and non-equivalence, 55–7, 385–7
  reaction time variances, 197–8
  self-terminating processing, 213

fast-guess model, 256, 266
fixed-point property, 266–9
fixed-sample-size model, 336, 339
Fourier transform, 34
functional equations, 38, 58 (Eqs. 4.10, 4.11), 397

gamma distribution, 39–40, 44, 46, 276, 285, 287
  general, 40–1, 209
geometric probability distribution, 43–4, 46
gestalt, 173, 421

hazard function, 26–7, 38, 45, 64, 248–54, 250–4
hybrid models, 61–5
hypergeometric distribution, 336

identity processing, 154–5, 416–17
independence, 28, 30, 39–40, 42, 45, 68–76, 324–5
  across-stage, 51–2, 69–71; stochastic, 361
  total completion time, of, 71–6, 320, 334–41
  whole report, in, 322–5, 336, 338, 348–9
  within-stage, 48–9, 52
information-processing approach, 6–7
intensity function, *see* hazard function
intercompletion time, 40, 49
inverse hypergeometric distribution, 174

joint distribution, 27–9, 45

latent network theory, 15–22
Laplace transform, 34–6, 41
late target paradigms, *see* memory scanning
level of processing, 65, 153–4
likelihood ratio, 281–2, 286, 294–7
linear systems, 401–9
long reaction time (RT) property, 220–7, 235–48
  empirical tests, 243–8
  statistical test, 242

marginal distribution, 28–9
masking, effects on serial and parallel accuracy, 325–9, 341–4
measurement scale of reaction time, 395–6
median processing time, 95–8
memory scanning or search, 115–26, 165–6, 190, 219
memoryless property, 38, 44, 46, 274
minimum completion time, 248–54
minimum reaction time, 128–30
models, and classes of, 49, 359, 448
moment-generating function, 32, 34, 36, 40, 44–6, 297–9
moments, 32, 36–8, 46
multicomponent model, 351–3

negative binomial distribution, 44, 46
nested states assumption, 218–32, 268
noisy operator theory, 155–6
non-Donderian response bias model, 145–8

optimal search model, 139–45

parallel processing, 42, 50–61, 83–98, 130–2, 228–31, 418–19
  independent: exhaustive, 171–2, 177–8, 191–2, 200, 209–10; first intercompletion time, 248–54; long-RT property, 235–7; self-terminating, 171, 191, 199–200, 213; short-RT property, 231–5
  limited capacity; independent, 85–8, 89–91, 93–7, 251–2; reallocation of capacity, 88–9
  linear target present and absent curves, 130–1
  supercapacity, 252–3
  unitary vs. multiple capacity sources, 131–2
  unlimited capacity: independent, 83–4, 92–3, 250–1; reallocation of capacity, 84–5
parallel vs. serial processing
  empirical testing, 121, 325–9, 383–7; manipulating search set variables, 157–62; parallel–serial testing paradigm (PST), 162, 414–47
  equivalence of models, 57–61, 457–62, 466, 468–70
  nonequivalence of models, 57–61, 463–6, 470–82
  see also additivity of reaction time
parallel–serial testing paradigm (PST), 162, 414–47
  distribution function level, generalized to, 445–7
  experimental application, 437–45
  simulated tests, 427–37
partial report paradigm, 317–19
personal equation, 4–5
Poisson distribution, 39–46, 62, 77–80, 274, 284–8, 322–3, 334–5
principle of correspondent change, 467–8, 481–2
processing order, 53, 64
probability space, 449–57
  model implication and equivalence, 452–7
pure insertion assumption, 219
pushdown stack model, 134–7

quantiles, 95–8

random walk models, 257–8, 291–316
  relative judgment theory, 293–7; unbiased model, 306–10
  SPRT model, 293–7, 302–5; biased, 310–14
reaction time (RT)
  history, 1–7
  measurement theory, 387–401
relative judgment theory, *see* random walk models
redundant target paradigms, 124, 157–62, 198–9
residual processing time, 120, 269
response execution, 118–19
response selection, 118–19, 135, 145–8

same–different paradigms, 153–7, 173
self-termination, 12–13, 65–8, 126–30, 164–205, 212–15, 218–48, 417
  increases in minimum reaction time, 128–30
  parallel target present and absent curves, 126–8
self-termination vs. exhaustive processing
  model equivalences and diversity, 60–8; *see also* Chapter 7
  empirical testing: accuracy, tests involving, 203–5; distribution tests, 201–3; equal slope test, 121–2, 166–8; experimental application, 179–82; manipulating search set variables, 168–79; serial position effects, 183–92; variance of reaction time, 124–5, 192–201
sequential probability ratio test (SPRT), 293
  see also random walk models
serial position effects
  accuracy, 321–4, 331–4
  reaction time, 122–4, 183–92
serial processing, 9–12, 39, 50–61, 80–3, 227–8, 418
  exhaustive, 189–90; density function, 208–11; different rates model, 123, 169–70; serial position effects, 189–90; standard serial exhaustive model, 80–2, 120–6, 190
  falsified by target present and absent curves, 148–51
  self-terminating, 184–9; density function, 212–13; fixed-path models, 184–9; parallel target present and absent curves, 167–8, 187–9; variable-path models, 184–7; variance predictions, 194–9
  see also parallel vs. serial processing; parallel serial-testing paradigm
short reaction time (RT) property, 221–7, 231–48
  empirical tests, 243–8
  statistical test, 242–3
short-term memory, 115, 117, 134–7
simultaneous memory and visual search, 151–3
speed–accuracy trade-off, 293
stage, 10, 358–61
stimulus encoding, 118
stochastic processing, task of modeling, 23–46
stochastic simultaneity, principle of, 472–5
subtraction method of Donders, 5, 358, 408, 410
survivor function, 27, 45
systems, 23–4, 358–9
  linear, 401–9
  realistic system differences: parallel vs serial, 53–4

target item, 115–16
target–nontarget rate differences, 154, 170–2
tight packing assumption, 218–31, 268
time-sharing models, 62–4
timing models, 257
total completion times, 49
trace strength models, 133
transmodal property, 284
two-alternative forced choice task, and reaction time and accuracy orderings, 258–60
two-state models, 260–71

variance of reaction time, 152, 264–5
  self-terminating vs. exhaustive testing, 124–5, 192–201
varied-state models, 255–6, 260–72
  multistate models, 271
  two-state models and empirical tests of, 260–71
visual search, 115–26, 165–6, 190, 219

waiting time, 77–80, 284–5
Wald's identity, 299–302
Weibull distribution, 159–62
whole report paradigm, 317–55
within-stage independence, 48–9, 52