The meaning of additive reaction-time effects: Tests of three alternatives.

Seth Roberts & Saul Sternberg

Reprinted from D. E. Meyer & S. Kornblum (Eds.) *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience - A Silver Jubilee.* Cambridge, MA : M.I.T. Press, 1993. Pp. 611-653.

26 The Meaning of Additive Reaction-Time Effects: Tests of Three Alternatives

Seth Roberts and Saul Sternberg

26.1 INTRODUCTION

At Attention and Performance II, Sternberg (1969) introduced the additive-factor method, for interpreting reaction-time (RT) data from factorial experiments. In that method, additivity of the effects of experimental treatments on mean RT is taken to suggest that the underlying mechanism can be divided into independently changeable, serially arranged operations (stage model). In this chapter we consider two other explanations of additive means: a model with independently changeable alternate pathways and the McClelland-Ashby cascade model. In all three models, experimental factors that influence different operations can have additive effects. To choose among these models we develop several tests, including comparisons of entire RT distributions. Applied to the results of four diverse experiments, the tests support the stage model and contradict the alternate-pathways and cascade models. In particular, the results of one distributional test, based on a suggestion by Ashby and Townsend (1980), support the stage model remarkably well.

Some of the examples of additivity in Sternberg (1969) were remarkably precise. In a numeral-naming experiment, for example, main effects of about 50 and 100 ms were additive with a precision of about 1 ms. Since then, many more examples of impressive additivity have been observed (Roberts 1987; Sanders 1980, 1990; Sternberg 1971). These results were unexpected; there is no obvious reason why behavior should be so simple. Nothing in the anatomy or physiology of the brain would lead one to expect such simplicity, and there is almost no precedent in other observations of behavior. Unlike the few other well-established cases where behavior has a simple quantitative structure, such as Stevens's power law, or receiver operating characteristics (ROCs) that are linear in normalized coordinates (e.g., Swets 1986), RT additivity is comparatively general, found in a large range of experiments, with diverse factors. Whatever the mechanism that produces additivity, it must be widespread.

What is the mechanism? The proposal that additivity reflects stages of processing has gained some support. Sanders (1980; 1990, especially fig. 1) has shown how existing examples of additivity, taken together and interpreted in terms of processing stages, make a consistent picture. Roberts (1987) showed that with some modification, the stages explanation of RT additivity

also explained multiplicative effects of factors on response rate in animal experiments, and that this explanation made sense in terms of other knowledge.

The proposal has also faced both specific challenges and general skepticism. As an example of a specific challenge, Pieters (1983) questioned the use of analysis of variance to test additivity. Other critics include Taylor (1976), Townsend (1984), Townsend and Ashby (1983), and Wickelgren (1977). For a summary, see Luce (1986, 481–483), who states that "many of those who have commented on the matter very strongly question the existence of stages at all, at least as conceived by Sternberg" (482). (Miller 1988 argues that some of the evidence regarded as negative is unpersuasive, however.) For examples of general skepticism see Broadbent (1984, 56–58) and Gardner (1985, 120–124). Such skepticism may reflect the belief that stage models are too simple; Broadbent, for example, calls them "simplistic" (55). But additivity is simple.

Alternative Explanations of Additive Effects

Regardless of current theories and beliefs, additivity as a phenomenon is too widespread to be ignored. Here we consider Sternberg's (1969) explanation together with two alternatives. The first, which produces exact additivity, is a mechanism with alternate pathways; the response is generated by one or the other pathway with a fixed mixing probability. The second, which produces approximate additivity, is McClelland's (1979) cascade model as further developed by Ashby (1982). See also Townsend and Ashby (1983, chap. 12).

Suppose that factors A and B have additive effects on mean RT. In all three of the explanations we consider, the mechanism is modular (Shallice 1988, sec. 2.1) in the sense that it is composed of processes *a* and *b* that can be changed independently. Additivity in an experiment with factors A and B is explained by the mechanism together with an assumption of selective influence: factor A influences process *a* but not *b*, whereas factor B influences process *b* but not *a*. Hence, for purposes of the present chapter, we incorporate an assumption of selective influence among the defining features of each model, from which we derive properties that permit it to be tested. The models differ in the arrangement of processes *a* and *b*, and in the nature of the communication between them.[1] In a *stage model* the processes are arranged sequentially, with one beginning when its predecessor is complete; both are required for response initiation. In the *alternate-pathways model*, process *a* is used on some occasions while process *b* is used on the remaining occasions; on no trials are they both used. In the *cascade model* both processes are used on all trials, but process *a* provides output continuously to process *b* ("partial output") which thus operates concurrently with *a*.[2]

Implications for the Additive-Factor Method

That radically different mechanisms are capable of producing additive effects of experimental manipulations on mean RT has important implications for the

additive-factor method. First, insofar as approximations to these mechanisms exist, it widens the domain within which the method can be used to discover and determine the properties of independently changeable (modular) processes.[3] But second, although a finding of additivity still supports such modularity, it does not support a stage model without additional evidence, contrary to Sternberg (1969). In what follows we provide such evidence, based on tests that distinguish among the three mechanisms.

26.2 GENERAL AND STOCHASTICALLY-INDEPENDENT STAGE MODELS

Definition of the Models

In the stage model, processes a and b operate in sequence, possibly concatenated with other operations; one process begins when its predecessor is complete (discrete transmission; see, e.g., Meyer, Yantis, Osman, and Smith 1984; Miller 1988, 1990; and Sanders 1990). For a description of three relations between processes that would produce such seriality, see Sternberg 1984. According to the model, the stream of processes between stimulus and response can be cut at some point, defining two processing stages, "stage a" the processes before the cut (including process a), and "stage b" the processes after the cut (including process b). The cut is temporal; it may or may not be spatial (anatomical). We make two assumptions, the first of which follows from the sequential arrangement of stages:

1. *Stages*: RT is the sum of stage durations: $T = T_a + T_b$.

2. *Selective Influence*: Factor A influences the duration of stage a, but not b; factor B influences the duration of stage b, but not a.

In other words, factor A acts only before the cut; factor B acts only after the cut. A third assumption strengthens the model considerably:

3. *Stochastic Independence (SI)*: Durations T_a and T_b are stochastically independent. We refer to a model with this property as an *SIstage model*, to distinguish it from a general stage model (Gstage model) for which assumption (3) is not made. Of course, any evidence that favors the SIstage model also supports a fortiori the Gstage model.

Additivity of Factor Effects on Mean and Variance

Suppose each factor can have two levels, indexed by $i = 1, 2$ for factor A and $j = 1, 2$ for factor B. Given selective influence we can then write T_{ai} for the duration of stage a when factor A is at level A_i, and T_{bj} for the duration of stage b when factor B is at level B_j, so that we have for the RT, T_{ij}, at levels A_i and B_j,

$$T_{ij} = T_{ai} + T_{bj}. \tag{1}$$

We shall assume that increasing an index i or j corresponds to increasing the

mean RT; we refer to such a change in level as "raising the level" of the factor. As we shall be considering the two-way interaction contrast of various quantities in what follows, it is convenient to define it symbolically: For any function or quantity S_{ij}, where $i = 1, 2$ and $j = 1, 2$ we let

$$I\{S_{ij}\} \equiv S_{11} - S_{12} - S_{21} + S_{22}. \tag{2}$$

The central property of the stage model is then[4]

$$I\{\mu_{ij}\} = 0. \tag{3}$$

To make similar statements about higher-order moments, we must know or assume something about the stochastic dependence of stage durations. For example, relations among the RT variances under the four conditions in a 2×2 experiment depend on the set of covariances $\gamma_{ij} \equiv \text{cov}(T_{ai}, T_{bj})$. If we invoke the SI assumption, then because this implies that all four covariances are zero, we have variance additivity:

$$I\{\sigma_{ij}^2\} = 0. \tag{4}$$

Equation (4) characterizes the stage model under conditions more general than stochastic independence of stage durations, however. By using the relation $\sigma_{ij}^2 = \sigma_{ai}^2 + \sigma_{bj}^2 + 2\gamma_{ij}$ we see that $I\{\sigma_{ij}^2\} = 2I\{\gamma_{ij}\}$. Equation (4) will thus be satisfied not only if the stage-duration covariances are all zero (required in the SIstage model), but also if they are nonzero but constant (which also implies $I\{\gamma_{ij}\} = 0$), or nonconstant but additive. The last two possibilities are implausible, however: let $D_a = T_{a2} - T_{a1}$ and $D_b = T_{b2} - T_{b1}$ be the increments in stage duration produced by raising the respective factor levels. Richard Schweickert pointed out to us that equation (4) (additive variances) obtains if and only if $\text{cov}(D_a, D_b) = 0$ (uncorrelated increments). It is implausible that we would have uncorrelated increments but correlated base durations, T_{a1}, T_{b1}. This observation leaves zero covariance of stage durations as the favored explanation of additive RT variance and renders such additivity even more important.

The Summation Test

Further implications of the SI property are expressed informally by

$$F_{T_{11}+T_{22}}(t) = F_{T_{12}+T_{21}}(t), \qquad (t \geq 0), \tag{5}$$

which asserts the stochastic equality of the two sums: $T_{11,22} \equiv T_{11} + T_{22}$ and $T_{12,21} \equiv T_{12} + T_{21}$, where $F_X(t) \equiv F_X$ is the cumulative distribution function (CDF) of the random variable X. The idea of directly comparing the two CDFs of equation (5) to test the SIstage model in a 2×2 experiment was first proposed by Ashby and Townsend (1980), who proved the equality of convolutions equivalent to equation (5): $F_{T_{11}} * f_{T_{22}} = F_{T_{12}} * f_{T_{21}}$, where the $\{f_{T_{ij}}\}$ and $\{F_{T_{ij}}\}$ are the density functions and CDFs, respectively, of the $\{T_{ij}\}$, and $*$ represents convolution. To use the relation in the way they advocate, however, requires estimating density functions (Silverman 1986) and then

performing numerical convolution. We prefer the simpler method of approximating the convolutions that is embodied in the *summation test*, and that is suggested by equation (5) and its constructive proof (see Appendix). Instead of numerical convolution, one creates samples from $T_{11} + T_{22}$ and $T_{12} + T_{21}$ by simply summing RTs from each of the two pairs of conditions, and then determines the empirical CDFs of the two sums.

Because the summation distributions $T_{11,22}$ and $T_{12,21}$ must be identical (except for sampling error) given the SIstage model, any arbitrary property of the distributions must be identical. Thus, one advantage over testing the additivity of cumulants of the four component distributions (a property theoretically equivalent to the distributional equality) is that measures of scale and shape of the summation distributions can be used that may be more stable, robust, and resistant than the sample cumulants (Mosteller and Tukey 1977; Ratcliff 1979), even if such measures are biased.

While sufficient for the summation test to work, stochastic independence of stage durations is not necessary. This is shown by the following example, for which we are indebted to Frank Norman: Let $T_{ai} = t_{ai} + z$ and $T_{bj} = t_{bj} - z$, and assume that t_{a1}, t_{a2}, t_{b1}, t_{b2}, and z are mutually independent. Then the summation test is satisfied, even though stage durations T_{ai} and T_{bj} are not independent. Here $\text{cov}(T_{ai}, T_{bj}) = \sigma_z^2$, an instance of nonzero but constant covariance that satisfies equation (4), as described above. Necessary conditions for the summation test—which must be stronger than the Gstage model, but weaker than the SIstage model—have yet to be discovered.[5]

Another sufficient condition for equation (5), noted by Frank Norman, can be described as a translation condition among the distributions of the $\{T_{ij}\}$, and does not require stochastic independence of the stage durations that contribute to T_{ij}. Under this condition the distributions differ by translation only; the distributions of the centered RTs, $t_{ij} \equiv T_{ij} - \mu_{ij}$ are therefore identical:

$$F_{t_{ij}} \equiv F_{T_{ij}-\mu_{ij}} = F, \qquad (i = 1, 2; j = 1, 2). \tag{6}$$

The translation condition is approximated by a stage mechanism in which the durations of stages a and b display little variability from trial to trial, so that most of the variability is contributed by stages other than those influenced by factors A and B; an increase in factor level then only adds a constant to the RT. (Although it differs in detail, we shall see that such a mechanism is similar in spirit to the cascade model, in which the units influenced by experimental factors are not inherently stochastic, so that the required variability is grafted onto a deterministic mechanism.) Given the translation condition, the summation test would not be particularly helpful, since it would require only means additivity (equation 3) in addition. For the test to be interesting, therefore, the RT distributions for different conditions should differ by more than just their means. In sections 26.7 and 26.8 we show this to be dramatically true for the data sets we consider.

26.3 THE ALTERNATE-PATHWAYS (AP) MODEL

Definition of the Model

In the second model we consider that generates additive RT effects, the task is accomplished by process *a* on a proportion *p* of the trials, and by process *b* on the remaining trials. We call the processes "pathways" to suggest the possibility of distinct anatomical substrates. We make three assumptions:

1. *Alternate Pathways*: With probability *p* the response is produced by pathway *a*; with probability $1 - p$ by pathway *b*.

2. *Selective Influence*: Factor A influences the duration of pathway *a*, but not *b*; factor B influences the duration of pathway *b*, but not *a*.

3. *Fixed Probability*: Neither A nor B influences the pathway probabilities, *p* and $1 - p$.

Generality: Embedded AP Structure Suppose the AP mechanism is preceded and followed by one or more other processes (stages *c* and *d*). The resulting mechanism can be regarded as an AP model with one pathway containing stages *c*, *a*, and *d*, and the other containing *c*, *b*, and *d*. For all properties of the AP model to apply, however, we must assume that factors A and B do not influence stages *c* or *d*.

Generality: Multiple Pathways Suppose multiple alternate pathways, with one subset $\{a_k\}$ of pathways influenced by factor A, a second subset $\{b_l\}$ by factor B, and a third subset $\{c_m\}$ influenced by neither A nor B. This is equivalent to a two-pathway model with pathway *a* a probability mixture of the $\{a_k\}$ and pathway *b* a probability mixture of the $\{b_l\}$ and the $\{c_m\}$. The important constraint is that no subset of pathways be influenced by *both* A and B. Thus, unlike the SIstage model, in which a third factor C that interacts with both A and B cannot be permitted to vary freely without inducing a spurious correlation between stage durations, such free variation in an AP model does not alter any of its properties; different levels of C can be regarded as corresponding to different members of pathway *a* and pathway *b* subsets.

Plausibility The AP model deserves serious consideration for several reasons. The alternate processing pathways of the model may correspond to different physical paths: the brain contains multiple anatomical pathways along the route from input to output. An argument from anatomy that has been used to support parallel and connectionist processes (Rumelhart and McClelland 1986) thus also lends plausibility to the AP model.[6] Different paths may correspond to different subsets of stimuli: In Atkinson and Juola's (1974) theory of memory recognition, for example, decisions about some items are based on familiarity while decisions about others require an extended search. According to some models of choice that have been successfully applied to data, RTs are a mixture of responses from different pathways, determined by

fluctuations in the subject's state: In the fast-guess model of Ollman (1966) and Yellott (1971), for example, some responses are stimulus-controlled, others guesses, and factors that influence the former do not influence the guessing probability, consistent with assumption 3. In an RT experiment with pigeons, Blough (1978; see also Luce 1986, sec. 6.3) found that on some trials the response was not controlled by the wavelength of the stimulus while on the others it was; again consistent with assumption 3, the proportion was unaffected by wavelength. The naming of printed words is believed to be accomplished by more than one route, not all requiring graphemic-phonemic conversion (Coltheart 1985). Assumption 3 would be plausible in this context if, for example, the choice of pathway were governed by a fixed attribute of the word, unknown to the experimenter, and independent of the factors manipulated explicitly. For a review of multiple-pathway models of human information processing, methods for their analysis, and supporting data, see Yantis, Meyer, and Smith (1991). They consider experiments in which the manipulations are believed to alter the pathway probabilities but have no effect on the pathways, complementary assumptions to those of the AP model.

The Central Property and the Mixture Test

Let $G_i(t)$ be the CDF of the duration T_{ai} when factor A is at level i, and let $H_j(t)$ be the CDF of the duration T_{bj} when factor B is at level j. With this notation the observed reaction time T_{ij} is a mixture of T_{ai} and T_{bj} with mixing probability p; the CDF is the weighted sum,

$$F_{ij}(t) = pG_i(t) + (1 - p)H_j(t), \qquad (t \geq 0). \tag{7}$$

This leads to the central property of the AP model

$$I\{F_{ij}(t)\} = 0, \qquad (t \geq 0), \tag{8}$$

which is equivalent to

$$\tfrac{1}{2}[F_{11}(t) + F_{22}(t)] = \tfrac{1}{2}[F_{12}(t) + F_{21}(t)], \qquad (t \geq 0); \tag{9}$$

we multiply by $\frac{1}{2}$ so that each side of equation (9) is a CDF. The power of equations (8) and (9) resides in their independence of the pathway probability, p. Because each side of equation (9) is the distribution of an *equal-probability mixture* of two populations of RTs, it can also be written

$$F_{\mathrm{mix}(T_{11}, T_{22})}(t) = F_{\mathrm{mix}(T_{12}, T_{21})}(t), \qquad (t \geq 0), \tag{10}$$

where mix(X, Y) denotes such a mixture of random variables X and Y; this equation can be contrasted with equation (5) in which each side is the distribution of the *sum* of the same RTs. Corresponding to the summation test for the stage model we thus have a *mixture test* for the AP model. If sample sizes are equal in paired conditions, the model asserts that pooling the RTs from conditions 11 and 22, and from conditions 12 and 21, should produce two samples with the same population distribution. If sample sizes are unequal, then comparison of the means of pairs of corresponding empirical CDFs may be preferable.

Two sets of implications of equation (8) lead to tests of special interest: one concerned with means and variances, useful where distributions are not available, the other concerned with relations among the distributions.

Additivity of Factor Effects on the Mean and Other Raw Moments

Unlike the stage model, the AP model constrains the relations among the means and higher-order moments across the conditions of a factorial experiment without added assumptions comparable to stochastic independence of stage durations. Because estimates of moments of increasing order are increasingly unstable, we focus on first and second moments. From equation (8) it is easy to show that for the raw moments of order r, $\{\mu'_{rij}\}$,

$$I\{\mu'_{rij}\} = 0, \qquad (r \geq 1). \tag{11}$$

When $r = 1$ we have additivity of means, $I\{\mu_{ij}\} = 0$, the property we are trying to explain. When $r = 2$ we have additivity of the second raw moments (not the variances):

$$I\{\mu'_{2ij}\} = 0. \tag{12}$$

Because $\mu'_2 = \sigma^2 + \mu^2$, equation (12) becomes

$$I\{\sigma^2_{ij}\} = -I\{\mu^2_{ij}\}. \tag{13}$$

It is convenient to introduce two notational conventions for the means in 2×2 experiments. First, let a dot subscript represent averaging over levels of the corresponding index; for example, $\mu_{.1} \equiv \frac{1}{2}(\mu_{11} + \mu_{21})$. Second, let a d subscript represent a difference between levels of the corresponding index, subtracting lower from higher; for example, $\mu_{d1} \equiv \mu_{21} - \mu_{11}$; we also define $\mu_{dd} \equiv \mu_{22} - \mu_{11} = \mu_{d.} + \mu_{.d}$. We choose indices for factor levels such that the main effects of both factors are nonnegative: $\mu_{d.} = \mu_{2.} - \mu_{1.} \geq 0$, and $\mu_{.d} = \mu_{.2} - \mu_{.1} \geq 0$. In what follows we assume that both main effects are nonzero. Given means additivity,

$$I\{\mu^2_{ij}\} = 2\mu_{d.}\mu_{.d}; \tag{14}$$

from equation (13) we then have

$$I\{\sigma^2_{ij}\} = -2\mu_{d.}\mu_{.d}. \tag{15}$$

Because $\mu_{d.}$ and $\mu_{.d}$ are both positive, the right-hand side of equation (15) is negative. The AP model thus requires that rather than being additive, as in equation (4), the effects of the two factors on the variance interact, with $2\mu_{d.}\mu_{.d}$ being the magnitude of the (negative) interaction contrast. If the main effects of A and B on the variance are in the same direction as their effects on the mean (often but not always the case; compare $na = 2$ and $na = 8$ in tables 26.2 and 26.3), then the interaction is negative (underadditivity).

The contrast between equations (4) and (15) implies that no set of distributions $\{F_{ij}\}$ can satisfy both the mixture and summation tests: For any distributions that satisfy the summation test, variance effects are additive (equation 4);

for any distributions that satisfy the mixture test, variance effects interact (equation 15). Furthermore, equation (15) indicates how the summation test should fail, given the AP model: Because $\sigma^2_{11} + \sigma^2_{22} < \sigma^2_{12} + \sigma^2_{21}$, $F_{T_{12}+T_{21}}$ will be flatter than $F_{T_{11}+T_{22}}$, and, to the extent that the two distributions are symmetric, the CDFs will cross close to their medians ($F \approx 0.5$). It is also of interest how the mixture test will fail, given a stage model. We need assume only variance additivity (equation 4), not full stochastic independence, to show that

$$\sigma^2_{\text{mix}\{T_{11},T_{22}\}} - \sigma^2_{\text{mix}\{T_{12},T_{21}\}} = 4\mu_{d\cdot}\mu_{\cdot d}. \tag{16}$$

Thus $F_{\text{mix}\{T_{12},T_{21}\}}$ will be steeper than $F_{\text{mix}\{T_{11},T_{22}\}}$; again, to the extent that the two distributions are symmetric, the CDFs will cross close to $F = 0.5$. The pattern expected in failures of the mixture test is further discussed below.

The expected failure of variance additivity produced by the AP model can be dramatic, sometimes a *crossover interaction*, in which the *direction* of the effect of one factor (not merely its magnitude) depends on the level of the other. Thus, let an "*i*-crossover" in a 2 × 2 experiment be an interaction in which the sign of the *i*-effect on a statistic S_{ij} depends on the *j* level, and let a "*j*-crossover" be defined similarly. It is easy to show that we have an *i*-crossover if and only if $|I\{S_{ij}\}| > 2S_{d\cdot}$ and a *j*-crossover if and only if $|I\{S_{ij}\}| > 2S_{\cdot d}$. For example, in experiment 2, $na = 8$, the mean predicted variance interaction contrast ($-2\mu_{d\cdot}\mu_{\cdot d}$) is $-10{,}824 \pm 1{,}894$ ms^2; this value can be compared to the main effects on RT variance (438 ± 166 ms^2 for factor A and 841 ± 297 ms^2 for factor B), after doubling them.[7] Both crossover conditions are satisfied; if the model is correct the variances must therefore display both kinds of crossover interaction.

The AP Model as a Special Stage Mechanism

That the AP model is a special case of the Gstage model (but different from the SIstage model) can be seen as follows: Suppose a stage model with stages *a* and *b* and corresponding factors A and B. Now suppose there is a third factor, C, that influences both stages so as to interact powerfully with factors A and B in a special way: At level C_1 of C the effect of A on *a* is nullified; at level C_2 of C the effect of B on *b* is nullified. If data are pooled over levels of C, we have an AP model in which the pathway on a trial is determined by C level. (The means additivity of the AP model thus follows from its being a stage model; the failure of variance additivity shows that this special stage model cannot be an SIstage model.) The AP and SIstage models are thus at the ends of a continuum of joint influence: Let an "A-influenced process" be a process whose duration depends on A. In the AP model there are no trials on which A-influenced and B-influenced processes both operate; in the SIstage model A-influenced and B-influenced processes operate on all trials. (Roberts, chap. 25, 600–601, describes two ways of distinguishing these possibilities in other situations.)

Tails of the RT Distribution and Mixture-Test Failures

It is helpful to consider how the mixture test might be expected to fail, given a stage model; we do so by supposing processes with minimum and maximum durations.[8] It is plausible for most psychological processes that their minimum durations are greater than zero: even the fastest output cannot be produced in arbitrarily short time. (Cf. Donders's ([1868] 1969) application of the subtraction method to the minimum RT.) With some important exceptions it also seems plausible that a change in factor level that increases the mean duration of a process also increases its minimum duration. (We shall see that neither of these properties applies to the cascade model, however.) In an SIstage model any change in factor A that increases $\min(T_a)$ will increase the minimum RT, $\min(T)$ to the same extent. (A corollary is that factors will have *additive* effects on the minimum, such that $I\{\min(T_{ij})\} = 0$, mirroring the additivity of μ_{ij}.) Finally, this property will be robust in the face of all except extreme forms of stochastic dependence (such as that embodied in the AP model), so we can expect it in many cases of the Gstage model. For the AP model, however, the property need not obtain, even though additivity in the mean does: Suppose that the shortest times are produced by pathway *a*. An increase in the level of factor B will then not change the minimum RT. In this respect the AP model acts like a (self-terminating) pair of parallel processes, in which the one with the shortest duration is reflected in the RT. In short, for a Gstage model it is possible (but not necessary) that increasing the level of *either* of two factors will increase the minimum RT; for an AP model *both* levels must be increased. A similar contrast between the models holds for the relation between maximum process durations and the maximum RT.

These intuitions about extreme RTs are captured by two corollaries of equation (8) that may focus analyses on aspects of the AP and stage models that are especially useful in model discrimination. The first is interesting in relation to short RTs:

$$F_{11}(t) \leq F_{12}(t) + F_{21}(t), \qquad (t \geq 0). \tag{17}$$

For small t, where $F_{22}(t)$ is small, equation (17) captures the intuition that, depending on which pathway contains the shortest times in condition 11, either F_{12} or F_{21} will contain RTs that are as short as the shortest in F_{11}. In contrast, increasing the level of either factor alone in a stage model can increase $\min(T)$; the short times in condition 11 would then be too abundant for equation (17) to be satisfied, and the left side of equation (9) would exceed the right, for small t. The second corollary is interesting in relation to long RTs. Let $\bar{F}_{ij}(t) \equiv 1 - F_{ij}(t) = Pr\{T_{ij} > t\}$. Then,

$$\bar{F}_{22}(t) \leq \bar{F}_{12}(t) + \bar{F}_{21}(t), \qquad (t \geq 0). \tag{18}$$

For large t, where $\bar{F}_{11}(t)$ is small, equation (18) captures the intuition that either F_{12} or F_{21} will contain RTs as long as the longest in F_{22}. In contrast, decreasing the level of either factor alone in a stage model can reduce $\max(T)$; the long times in condition 22 would then be too abundant for equation (18) to be satisfied, and the right side of equation (9) would exceed the left, for large t.

This pattern of mixture test failures expected for some stage models corresponds to what we expect from the behavior of the variance of the SIstage model, discussed above.

26.4 THE MCCLELLAND-ASHBY CASCADE MODEL

Definition of the Model

The third model we consider, the cascade model, was introduced by McClelland (1979) and further developed by Ashby (1982). Even assuming selective influence of factors on processes, it does not invariably generate additive RT effects, but with some parameter settings additivity is approximated well. This model is of interest partly because it is a precursor of more complex connectionist models, which often incorporate some of its features (Rumelhart and McClelland 1982).

The model contains a set of $k = 1, 2, \ldots, n$ processing units, each with an input and an output. The output from unit k is the input to unit $k + 1$. The amount of its work accomplished by unit k by a given time t after stimulus onset is measured by its level of output activation $a_k(t)$ at that time; this grows continuously and at a rate proportional to the instantaneous difference between its current input and output levels (a linear integrator), divided by its time constant, τ_k. Thus, in response to a step-function input at time $t = 0$ the output grows exponentially to an asymptote, α: $a_k(t) = \alpha(1 - e^{-t/\tau_k})$. Stimulus onset provides such a step-function input to unit 1. The response mechanism is triggered when the output of unit n exceeds an activation criterion. RTs vary from trial to trial because of variability in the time added by the response mechanism, and because of noise in the final output activation level or, equivalently, in the criterion; the units themselves function deterministically.[9]

In the model's initial formulation, experimental manipulations could influence the asymptotic activation level as well as the time constants. Because of this feature, as well as the choice of noise distribution, there could be too many trials on which the criterion exceeded the asymptotic activation level, precluding a response. To eliminate this problem, Ashby (1982) permitted experimental factors to influence only the time constants, and adjusted the noise distribution slightly, truncating it at $+2.5\sigma$, so that the criterion never exceeds the asymptotic activation level. With these adjustments and a few others Ashby derived the RT CDF that we have used in our explorations.

The cascade model is much more specific than the stage and AP models; for example, it does not share their virtue of being distribution free. It is not obvious how best to characterize its many features, nor is it yet clear which features are essential for its interesting properties. For example, it is not known how important are the particular shape, location, and spread of the criterion distribution, nor the particular law for the growth of activation. On the other hand, not all such laws are consistent with the spirit of the model, since some could transmute it into a stage model. One consequence of our current ignorance about which features are critical is that evidence against the model, which may result from an incorrect choice of particular features or parameter

values, is relatively less important than evidence for it, which is unlikely to result from a combination of such incorrect choices. By testing this particular model we hope to learn one way to approach the more interesting goal of testing a larger class of models of which it is a member. One possible description of the model is as follows; the first four assumptions refer to quantitative features that must be specified precisely.

1. *Processes in Cascade*: Processing is gradual, with the current results of one process immediately available to the next.

2. *Unit Time Course*: A linear integrator relates $a_k(t)$ to $a_{k-1}(t)$, as described above.

3. *Response Actuation*: A response is triggered when $a_n(t)$ reaches criterion.

4. *Noise Distribution*: Here is specified the criterion (or activation) noise distribution and the assumption that the mechanism is affected by just one sample from this distribution per trial.

5. *Selective Influence*: Factor A influences only process a; factor B influences only process b.

6. *Influence Mechanism*: A factor influences a process by altering its time constant, τ_k.

The processing units in the cascade model are ordered structurally by their input-output relations. Only in some senses are they ordered temporally. For example, a given level of activation must be achieved by unit k before it can be achieved by unit $k + 1$, but activation begins to rise at the output of the final unit as soon as the stimulus is applied, and responses can be triggered before any process is close to its asymptote. Furthermore, the output activation function $a_n(t)$ is independent of unit order, evident in equation (19). In addition to its relevance to connectionist models, the cascade model is interesting because, as we shall see, it is capable of producing approximately additive factor effects on both mean and variance, like the SIstage model, and because it embodies an interesting and plausible idea—that information is passed continuously from one process to the next. The idea of "partial output," of which this is one realization, has been tested (with mixed conclusions) by devising special experimental procedures or using measures other than RT; see, for example, Osman, Bashore, Coles, Donchin, and Meyer (1992), Meyer, Irwin, Osman, and Kounios (1988), Meyer, Osman, Irwin, and Yantis (1988), Miller (1988), Miller and Hackley (in press), and Schweickert (1989). Testing of the cascade model is an additional approach.

The Processing-Time Distribution

Activation level at the output of unit n, as a proportion of the asymptotic value, behaves according to the cascade equation:

$$E_n(t) = 1 - \sum_{k=1}^{n} \left[\prod_{\substack{m=1 \\ m \neq k}}^{n} \frac{\rho_m}{\rho_m - \rho_k} \right] e^{-\rho_k t}, \qquad (0 \leq t < \infty, n \geq 2), \tag{19}$$

where $\rho_k = \tau_k^{-1}$ is the rate constant of unit k. As t grows from 0 to ∞, $E_n(t)$ grows from 0 to 1. Time, t, is in arbitrary units; to connect the model to data requires either using statistics in which the time unit is eliminated, or specifying t_u ms, the duration of a time unit, and multiplying t in equations (19) and (20) by t_u. Following Ashby (1982) we set the asymptotic activation level at 5 units and the mode of the adjusted criterion distribution at 2.5 units; the distribution of the time to reach criterion is then:

$$G_n(t) = \frac{\Phi(5E_n(t) - 2.5)}{\Phi(2.5)}, \qquad (0 \leq t < \infty), \tag{20}$$

where $\Phi(\cdot)$ is the standard normal CDF. The numerator is $\Phi(-2.5)$ when $t = 0$, and approaches $\Phi(2.5)$, or about 0.994, for large t; the term in the denominator normalizes the truncated distribution.

Additivity of Factor Effects on the Mean and Variance

Given a specification of the number of units and their time constants, one can use the CDF, $G_n(t)$, to compute any statistic one wishes. Sufficiently little is known about the model, however, so that one must perform such computations over a subspace of parameters, rather than determining statistics analytically as functions of the set of time constants, or independently of them. One consequence is that the generality of our conclusions may be limited. In considering the implications of the computed statistics for observed RTs one must also keep in mind that the assumed cascade process may be concatenated with other components, such as other cascade units or processing stages.

To investigate the model, we had to decide on a plausible range of values for the set of time constants τ_a and τ_b that we manipulated. Given that the time unit, t_u is free to vary, only the *ratios* of time constants matter, so no loss of generality results from specifying one arbitrarily; we thus defined 1.0 time unit as the largest value. We chose 20:1 as the maximum ratio of time constants to examine, as did Ashby (1982), so that in most calculations, τ ranged downward to about .05. (In supplementary calculations we worked with ratios of time constants up to 100:1, and found that our conclusions were not altered, as will be seen in section 26.10.) Some considerations that might justify the choice of 20:1 are as follows: (a) The contribution of a cascade unit to $\overline{RT}$ is proportional to its time constant. Even if the highest factor level (longest time constant, i.e., 1.0) is associated with as much as 300 ms, the lowest level would then be associated with only 15 ms. Given what we know about elementary cerebral events it seems unlikely that anything one would call the same process could take both as much as 300 and as little as 15 ms. (b) A unit is eliminated from the model when its time constant is set to zero. The time constant associated with the lowest attainable factor level may thus be small, but cannot be zero, as we don't wish to permit the lowest attainable level of a factor to entirely eliminate the process it influences. (c) It is likely that the lowest level of a factor in an actual experiment is higher than the lowest attainable level of that factor.

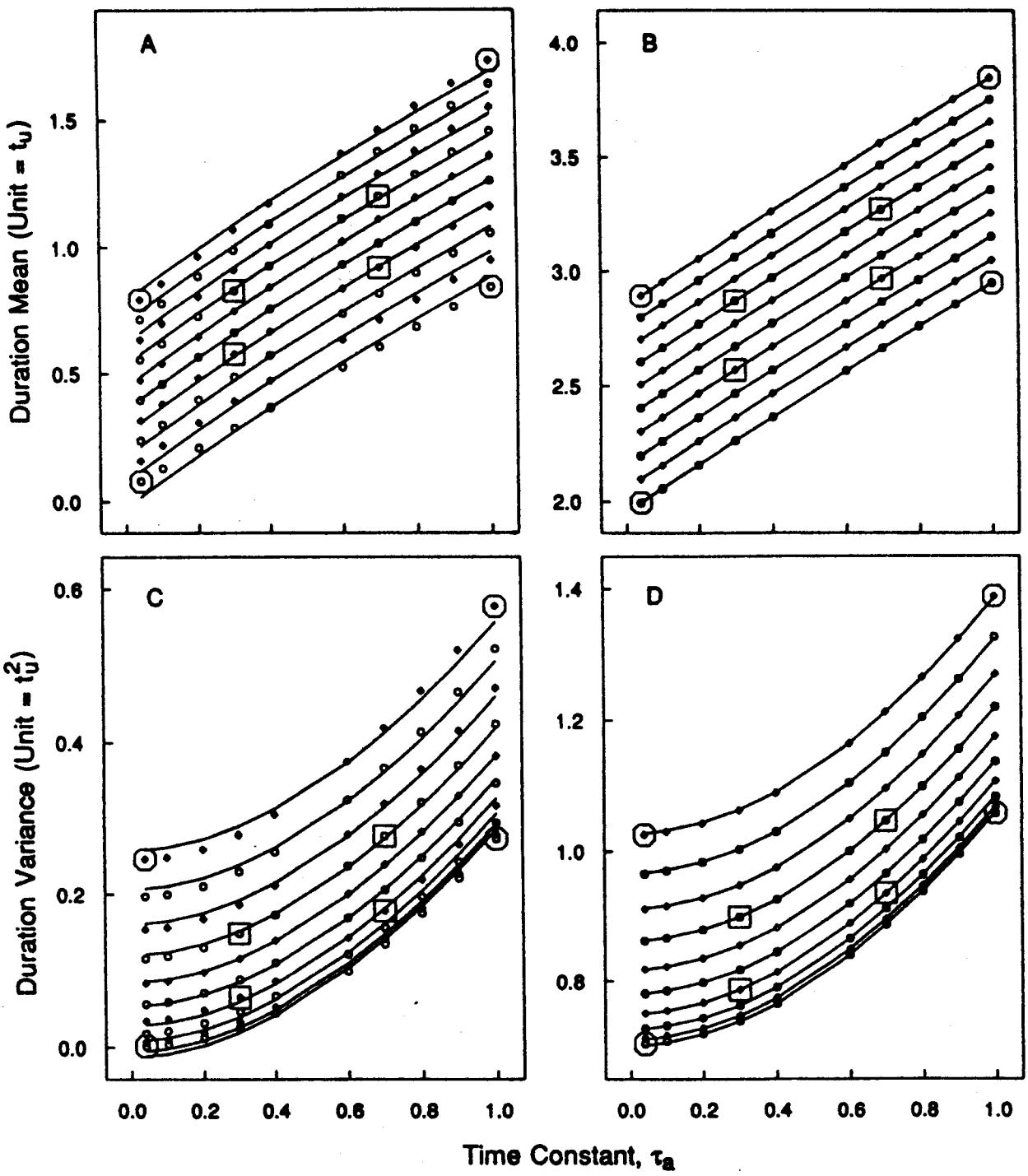

Figure 26.1 Behavior of the mean and variance in two cascade models (see table 26.1). In each panel the statistic is shown as a set of points plotted as a function of τ_a, with τ_b as the parameter distinguishing the sets of points. The values of τ_a are 0.04, 0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.9, and 1.0; the ten increasing values for each τ_a are generated by $\tau_b = .05, .15, \ldots, .95$. The ten curves in each panel are parallel, the least squares fit of an additive model to the set of 100 points. *Panel A:* Means for model 2.1 in arbitrary units, t_u. *Panel B:* Means for model 4.3. *Panel C:* Variances for model 2.1 in arbitrary units, t_u^2. *Panel D:* Variances for model 4.3. The subsets of points marked by large squares and large hexagons indicate values used for the interaction measures given in table 26.1.

Table 26.1 Interaction of Effects on Mean and Variance in Six Cascade Models

Model	τ_c	τ_d	$\frac{100\, I\{\mu_{ij}\}}{\sqrt{\mu_{d\cdot}\,\mu_{\cdot d}}}$ Large Effects	$\frac{100\, I\{\mu_{ij}\}}{\sqrt{\mu_{d\cdot}\,\mu_{\cdot d}}}$ Small Effects	$\frac{100\, I\{\sigma^2_{ij}\}}{\sqrt{\sigma^2_{d\cdot}\,\sigma^2_{\cdot d}}}$ Large Effects	$\frac{100\, I\{\sigma^2_{ij}\}}{\sqrt{\sigma^2_{d\cdot}\,\sigma^2_{\cdot d}}}$ Small Effects
2	0	0	21.5	10.0	20.6	13.8
3.1	0.32	0	11.6	7.2	17.6	11.4
3.2	1.05	0	0.4	0.8	3.0	0.0
4.1	0.32	0.73	3.4	2.3	7.7	3.3
4.2	1.05	0.32	1.1	0.8	3.4	0.5
4.3	1.05	1.10	−0.1	−0.1	2.9	0.1

For each model are shown values of the time constants τ_c and τ_d of units uninfluenced by factors A and B. (A zero time constant is equivalent to absence of the corresponding unit; the first digit of the model's designation is the number of units with nonzero time constants.) The interactions are those obtained by orthogonal combination of $\tau_a = 0.04, 1.0$ with $\tau_b = 0.05, 0.95$ (large effects), and $\tau_a = 0.3, 0.7$ with $\tau_b = 0.35, 0.65$ (small effects). Corresponding points for two of the models are emphasized in figure 26.1. The interaction measure is the interaction contrast as a percentage of the geometric mean of the main effects. (Normalizing in this way has little theoretical justification and, because main effects are similar across models, does not alter the relative sizes of interactions. Because the measure is dimensionless, however, it permits comparison of interactions of effects on means and variances, and also permits comparisons to data.) The models illustrated in figure 26.1 are models 2 and 4.3.

In figure 26.1 are shown the means and variances generated in two factorial simulations in which τ_a and τ_b were varied over the one hundred points in the orthogonal grid described in the caption.[10] Panels A and B show the means we obtained (points) together with a fitted additive model (parallel curves). Results in panel A, from a two-process cascade, show an overadditive (positive) interaction. Consistent with earlier findings, the results in panel B, from a four-process cascade in which processes *a* and *b* were joined by fixed slow processes *c* ($\tau_c = 1.05$) and *d* ($\tau_d = 1.1$), show a remarkably good approximation to additivity. In panels C and D are shown the variances for the same pair of models. Again the two-process model shows overadditivity, while effects in the four-process model are remarkably additive. Table 26.1 provides descriptions of six different cascade models, including those in figure 26.1, together with measures of $I\{\mu_{ij}\}$ and $I\{\sigma^2_{ij}\}$. Confirming McClelland's (1979) suggestion for means, the interaction for both means and variances is negligible if the model contains a unit with a long time constant that is unaffected by factors A or B.[11] When fixed slow processing units are incorporated in a cascade model, the pattern of factor effects on both the mean and variance of the processing time can thus mimic the pattern produced by a stage model with uncorrelated stage durations, a remarkable feature of the cascade model given its dependence, in the stage model, on the RT being a sum of stage durations, and given that units in a cascade model operate concurrently.[12]

Figure 26.1 also shows that the increase in mean processing duration with τ is approximately linear, and that as the level of a factor increases, the variance considered as a function of the mean will accelerate. To test this property requires a one-factor experiment that is precise (because the contrast incorporates three sample variances and three means) in which performance is measured at three or more levels of that factor.[13]

26.5 FOUR EXPERIMENTS

We tested the three models with five sets of RTs for correct responses from four experiments, selected because they produced additive effects of two factors on mean RT, and because RT distributions, or at least variances, were available for individual subjects and for each combination of factor levels. Factors we shall call "A" and "B" are indicated below. Where factors had more than two levels, we selected pairs of levels (four conditions) for 2 × 2 analyses; where not otherwise stated, analyses depend on only those pairs for which mean RTs differed the most ("corners" of the design). For experiments 1 and 2 individual trials data were available, permitting distributional tests: for experiments 3 and 4 we used only means and variances. Sizes n of data sets are given as approximate mean observations per subject per condition. For experiments 1 and 2 we tried to reduce effects of heterogeneity by performing computations (such as moment calculations) within subsets of the data within each condition, and then averaging the results. The factor used to partition the data to form such subsets is indicated.

Experiment 1: Detection

Seven subjects produced simple reaction times, responding to a flash at one of four locations by pulling a lever (Backus and Sternberg 1988, experiment 1). Stimulus probability was high (.6) in one location and low (.05) in each of the other three; the remaining probability (.25) represents catch trials. The high-probability location changed from trial to trial, and was indicated by a central visual cue. The four locations were corners of an imaginary rectangle centered on a fixation point and separated from it by about 9 degrees of visual angle. The factors we consider are *foreperiod* (A) and *intensity* (B): the interval from warning signal to flash was one of six values: 750, 950, 1150, 1350, 1550, or 1750 ms; flash intensity was high, medium, or low. Levels of both factors varied randomly from trial to trial. Data were collected in six one-hour sessions after three hours of practice. For the present tests we restricted our analysis to responses to flashes in high-probability locations, and to trials on which the intensity was either high or low. The effect of foreperiod on mean RT is U-shaped, and is more convincingly additive with intensity for short than long foreperiods; we therefore used 750, 950, and 1150 ms as levels of A, and high and low as levels of B. We excluded data for one subject, for whom the main effect of A was markedly and significantly smaller than its effect on the other six subjects, but with no effect on conclusions. Calculations were done sepa-

rately for each session (there were six sessions) and then averaged over sessions; $n \approx 96$.

Experiment 2: Identification

Five subjects saw numerals and responded with spoken digits (Sternberg 1969, experiment V). The number of alternative stimulus-response pairs (*na*) was either 2 or 8. Within each of these there was a 2 × 2 factorial design, with factors *stimulus quality* (intact, or degraded by a superimposed checkerboard pattern; A) and *S-R compatibility* (name x, name $x + 1$; B). Data were collected from each subject in seven one-hour sessions, after five hours of practice. We have conducted separate tests for each of the two levels of *na*. We rejected seven of the observations (0.2%) for $na = 2$ and 14 (0.3%) for $na = 8$ as outliers, because each differed by more than 4 *SD*s from the mean of the remaining observations for that subject and condition. For $na = 2$ we reversed the sense of the compatibility factor for one subject whose main effect was opposite to the other subjects. Because of changes in analysis, means and variances differ slightly from those reported by Sternberg (1969). For $na = 2$, calculations were done separately for each numeral-repetition combination (two numerals × repeat/nonrepeat of the prior stimulus), then averaged over the four combinations; $n \approx 204$. For $na = 8$, calculations were done separately for each numeral (eight numerals), then averaged over numerals; there were no immediate repeats; $n \approx 245$.

Experiment 3: Classification

Twelve subjects served in an item-recognition experiment (Sternberg 1967). The stimulus was a numeral, and the response pulling a lever. One lever ("positive") was correct if the stimulus was contained in a memorized set of one, two, or four elements; otherwise the "negative" lever was correct. Positive responses were required on about 27% of the trials. Factors were *stimulus quality* (intact, or degraded by a superimposed checkerboard pattern, A) and *set size* (B). The analyses below are based on only the data from the more frequent negative responses, and only from the second of two sessions, during which an interaction between A and B present during the first session had disappeared; $n \approx 22$.

Experiment 4: Overlapping Tasks

Twenty-two subjects performed two binary-choice tasks on each trial (McCann and Johnston 1989). Task 1 was pitch discrimination: The stimulus was a tone of one of two frequencies, with "high" the correct response for the high tone, and "low" for the low tone. Task 2 was size discrimination: the stimulus was a rectangle of one of four sizes; the correct response was a button press with one hand for "very small" (S1) and "small" (S2), and with the other hand for "large" (S3) and "very large" (S4). We analyze the RTs in task 2.

Table 26.2 Statistics Based on Means in Four Experiments

Experiment	$m_{..}$	$m_{a.}$	$m_{.a}$	$I\{m_{ij}\} \pm SE_b(SE_p)$
1	222	15	36	-0.6 ± 2.1 (1.9)
2 ($na = 2$)	354	30	21	3.2 ± 3.7 (4.5)
2 ($na = 8$)	449	53	102	1.4 ± 3.8 (2.7)
3	458	69	121	4.5 ± 12.2
4	648	264	59	-4.0 ± 14.7

The unit is 1 ms. Data for each experiment are from a 2 × 2 set of conditions. $m_{..}$ represents the overall mean RT, $m_{a.}$ the main effect of factor A on the mean, and $m_{.a}$ the main effect of factor B. All main effects are statistically significant across subjects. $I\{m_{ij}\}$ is the interaction contrast of means (see equation 3). SE_b is based on between-subject variation; where available we also show SE_p, based on variation pooled over data subsets and subjects. For experiment 1, with six subjects, each with six data subsets, SE_b is based on 5 *df* and SE_p is based on 35 *df*. Corresponding degrees of freedom for experiment 2 are 4 *df* and 15 *df* for $na = 2$, and 4 *df* and 39 *df* for $na = 8$.

Factors were *stimulus onset asynchrony* (SOA; the time between onsets of tone and rectangle: 50, 150, 300, or 800 ms; A) and *discriminability* (the closeness of the rectangle to the classification boundary: near, for S2 and S3, or far, for S1 and S4; B). Levels of both factors varied randomly from trial to trial. Data are from the final 384 trials of the single session, following 128 practice trials. Trials were excluded if (a) either response was an error, (b) $RT_1 > 1000$ ms, (c) $RT_2 > 1500$ ms, or (d) either RT departed from its cell mean by more than 3 SDs. For these reasons 2.5% of the trials were excluded; $n \approx 47$. See Pashler and Johnston (1989) for discussion of overlapping tasks experiments.

26.6 ADDITIVITY OF FACTOR EFFECTS ON THE MEAN

In most of our tests we replace theoretical quantities by corresponding sample estimates, and base our estimates of precision on differences between subjects. Here we provide tests of the property shared by all three models, $I\{\mu_{ij}\} = 0$, by evaluating $I\{m_{ij}\}$; results are shown in table 26.2, together with the overall mean RT for the four experimental conditions considered, and the mean sizes of the two main effects. The interactions are all small compared to the main effects, and nonsignificant.[14] SE_b is sufficiently close to SE_p so as to indicate no important individual differences in $I\{m_{ij}\}$.

26.7 RELATIONS AMONG THE VARIANCES

Additivity of Factor Effects

Here we provide tests of variance additivity, $I\{\sigma_{ij}^2\} = 0$, which is expected from the SIstage model and weaker variants, and is well approximated by some variants of the cascade model. Results of evaluating $I\{s_{ij}^2\}$ are shown in the

Table 26.3 Statistics Based on Variances in Four Experiments

Experiment	$s^2_{..}$	$s^2_{a.}$	$s^2_{.d}$	$I\{s^2_{ij}\} \pm SE_b(SE_p)$	$I\{s^2_{ij}\} + 2m_{a.}m_{.d} \pm SE_b$
1	6.1	2.8	3.1	-1.2 ± 2.2 (2.0)	10.1 ± 3.0 ($p = .02$)
2 ($na = 2$)	26	9.2	4.0^{ns}	-0.2 ± 3.5 (3.8)	14.7 ± 2.0 ($p = .002$)
2 ($na = 8$)	13	4.4	8.4	1.8 ± 0.8 (1.4)	116 ± 21 ($p = .005$)
3	57	21^{ns}	53	18 ± 27	180 ± 47 ($p = .003$)
4	258	140	42	-40 ± 52	279 ± 69 ($p = .0005$)

The unit is 100 ms^2. Data for each experiment are from a 2 × 2 set of conditions. Where possible (experiments 1 and 2), we calculated variances for subsets of the data and then averaged within subjects. $s^2_{..}$ represents the overall mean variance of the RT, $s^2_{a.}$ the main effect of factor A on the variance, and $s^2_{.d}$ the main effect of factor B. All main effects are statistically significant except those with the superscript *ns*. $I\{s^2_{ij}\}$ is the interaction contrast of variances (see equation 4). $I\{s^2_{ij}\} + 2m_{a.}m_{.d}$ is the variance contrast whose expectation is zero, given the AP model (see equation 15). See table 26.2 for definitions of SE_b and SE_p. The *p*-values in the final column are based on two-tailed *t*-tests of the AP variance contrast versus zero; the *df* is one less than the number of subjects.

fifth column of table 26.3. Also shown are basic variance data for the four experiments: mean variances over the four conditions considered, and mean sizes of the main effects of factors A and B on the variance. The interaction contrasts are uniformly nonsignificant. SE_b and SE_p tend to be large, however—often as large as the smaller of the two main effects. Although these data are consistent with variance additivity, they are less convincing than the means additivity results.

Evidence against the Alternate-Pathways Model

Results of the variance test for the AP model are shown for all five data sets in the final column of table 26.3. Because $m_{a.}$ and $m_{.d}$ are orthogonal, we can assume that $E(m_{a.}m_{.d}) = \mu_{a.}\mu_{.d}$; we are therefore justified in substituting sample moments for theoretical quantities in equation (15). The differences between the two sides of that equation, which should be zero, are substantially and significantly positive in every case. For all the data sets except experiment 2, $na = 2$, the expected interaction contrast is so large relative to the main effects of both factors on the variance that both an *i*-crossover and *j*-crossover are required.

26.8 THE SUMMATION TEST: FURTHER SUPPORT FOR THE STAGE MODEL

We regard the summation test and the remarkable support it provides for the stage model as the most important contribution of the present chapter; we therefore describe the test and results in some detail for the three data sets to which we could apply it.

Violation of the Translation Condition

As discussed in section 26.2, if the translation condition (equation 6) were satisfied by the T_{ij}, then the summation test would add nothing to means additivity. Conversely, given means additivity, the degree to which one is impressed by success of the summation test should increase with the extent to which the T_{ij} distributions differ in more than location. That there are significant main effects on the variance (table 26.3) is one indication that the condition is violated, but the failure is more pervasive: in each of the three data sets, each of the second, third, and fourth moments about the mean, μ_2, μ_3, and μ_4, increases from condition 11 to condition 22 for every subject. The effect is significant (by two-tailed t-test) for all four moments in experiment 1 and experiment 2, $na = 2$. For experiment 2, $na = 8$, there is sufficient variation in the magnitude of the increase of the third and fourth moments so that the same test produces p-values of $p = .05$ and $p = .08$, respectively. The increases are also large: over the three data sets, $\mu_2^{1/2}$ increases by an average factor of 1.5, $\mu_3^{1/3}$ increases by an average factor of 2.1, and $\mu_4^{1/4}$ increases by an average factor of 1.7.

The Test Procedure

Partitioning of Data Suppose that factor A influences stage a, and factor B influences stage b, and that the SIstage model is valid. If the data include a mixture of levels of a third factor, C, that influences both stages (as level of practice or particular stimulus-response pair might), then this can induce a nonzero covariance of stage durations and cause the test to fail. If the level of C on each trial is known, however, the data can be partitioned into subsets within which the level of C is fixed, and the test applied separately to each subset, eliminating the problem, at least for that factor. This is the approach we took: In the $na = 8$ condition of experiment 2, for example, we noted that stimulus numeral had a systematic effect on mean RT, and used it as the partitioning factor, performing the test separately for the trials involving each of the eight numerals, in each of the four conditions within each subject. An incidental advantage of this approach is that it provided eight separate summation-test comparisons per subject, permitting within-subject measures of precision and tests of significance.

The Cartesian-Product Sums The summation test requires estimation of the two distributions, $T_{11,22}$ and $T_{12,21}$. Consider the first of these, for example. To obtain an estimate we created a sample by forming the Cartesian product of the sets of observations in conditions 11 and 22, and, for each pair in the product, determining the sum of its two members.[15] Thus in experiment 2, $na = 8$, an observation set in each of the four conditions for a particular numeral and subject contained about 31 RTs; the Cartesian-product sum that provided the sample of $T_{11,22}$ for that numeral and subject thus contained about $31^2 = 961$ values,[16] as did the sample of $T_{12,21}$ for that numeral and

subject. We use $\hat{F}_{11,22}(t, k)$ and $\hat{F}_{12,21}(t, k)$, respectively, to denote the empirical CDFs of these *summation sets* (*summation CDFs*), where k ranges over the data subsets in an analysis; thus, for eight subsets per subject and five subjects, $k = 1, 2, \ldots, 40$.

Adjustment of Distributions For each pair of summation CDFs we computed several statistics to be compared. However, before doing so we adjusted the locations and scales of each pair. To make statements about the data for each subject separately we planned to average the computed statistics over data subsets within subjects and use subset differences as a basis for variability estimates. To make statements about the group data and inferences about the population of subjects, we planned to average the computed statistics over subjects as well, and to base variability estimates on subject differences. The adjustment prior to averaging is based on the idea that any systematic failures of the summation test are more likely to occur at points with equal p-values in different pairs of distributions, than, for example, at points with equal time values. Another reason for adjustment was graphical: to increase the similarity of the shape of the average distribution to the shapes of the distributions being averaged.

For each pair of summation distributions we let a location parameter λ_k be the mean of their medians, and a scale parameter ξ_k be the mean of their interquartile ranges. We then determined the means, $\lambda_\cdot$ and $\xi_\cdot$ of these parameters over all data sets and performed the same linear transformation on all the values (X) in the kth pair of summation distributions:

$$X^* = \lambda_\cdot + \left(\frac{\xi_\cdot}{\xi_k}\right)(X - \lambda_k) \tag{21}$$

The result is that $\lambda_k^* = \lambda_\cdot$ and $\xi_k^* = \xi_\cdot$ for all k. For graphical purposes it was important to perform a corresponding adjustment to create the transformed component distributions T_{11}^*, T_{12}^*, T_{21}^*, and T_{22}^*, such that the summation property would be preserved (e.g., $T_{11,22}^* = T_{11}^* + T_{22}^*$), as well as the relative differences among the locations of the components. We accomplished this by using the same transformation (equation 21), but with $\lambda_\cdot$ and λ_k replaced by $\frac{1}{2}\lambda_\cdot$ and $\frac{1}{2}\lambda_k$, respectively.

Results of the Test

After forming the four adjusted component CDFs, $\{\hat{F}_{ij}^*(t, k)\}$, and the two adjusted summation CDFs, $\hat{F}_{11,22}^*(t, k)$ and $\hat{F}_{12,21}^*(t, k)$, for each data subset, we computed their means over data subsets to obtain CDFs for individual subjects. Means of the resulting CDFs over subjects are shown in figure 26.2. In all three data sets, the agreement between the two summation CDFs is remarkably good.

It is not obvious in which way the summation test will fail, if it fails. We therefore examined three different sets of statistics of the adjusted distributions in each pair, $\hat{F}_{11,22}^*(t, k)$ and $\hat{F}_{12,21}^*(t, k)$: proportions; quantiles and quantile-

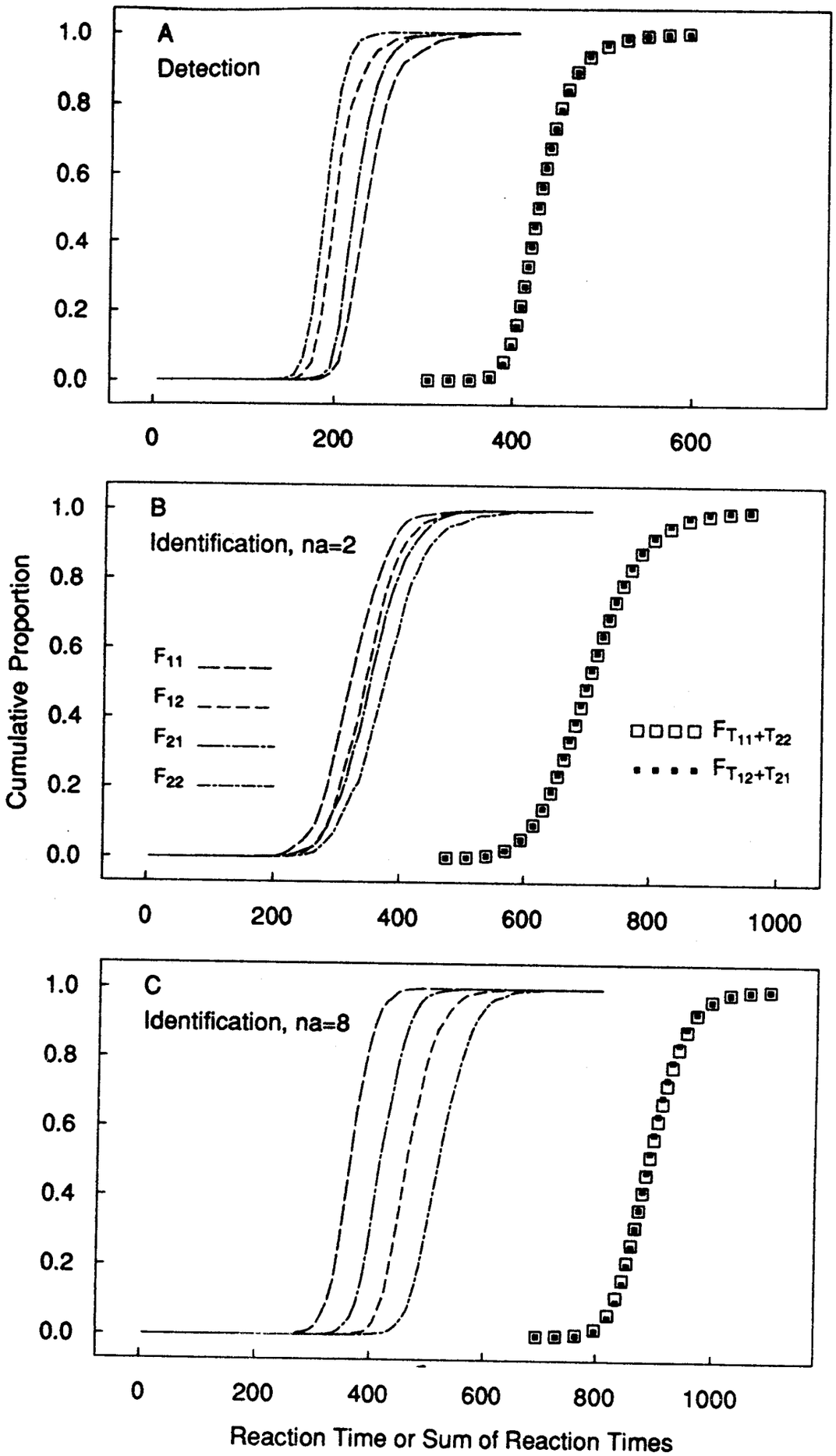

Figure 26.2 Results of the summation test for experiment 1 (panel A) and experiment 2 (panels B, C). CDFs were adjusted and averaged as described in the text. At the left of each panel are the average CDFs for each of the four conditions. At the right of each panel are the two summation CDFs; to enhance the visibility of the small differences we use symbols instead of curves. The scaling of the x-axis varies from panel to panel.

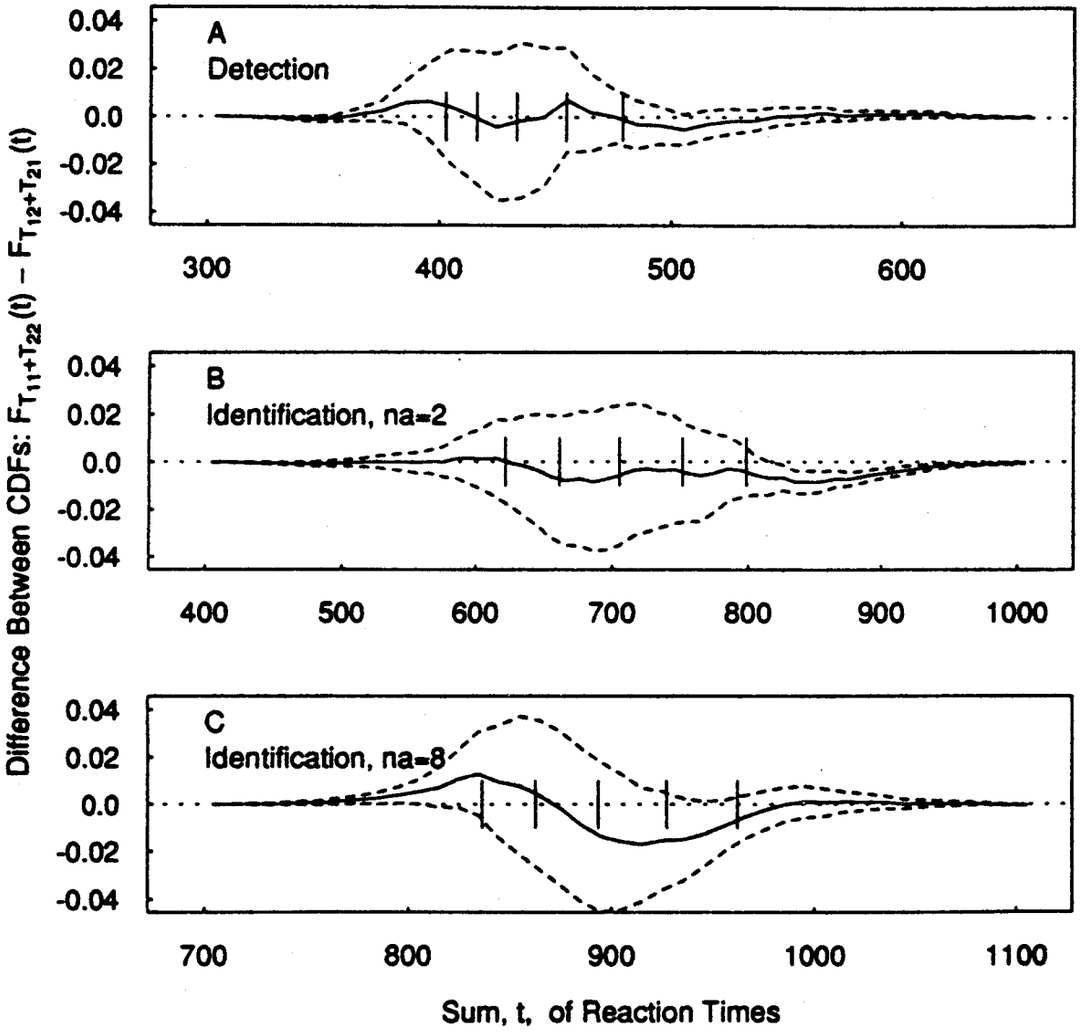

Figure 26.3 Deviations from CDF equality in the summation test for experiment 1 (panel A) and experiment 2 (panels B, C). On a greatly enlarged ordinate scale, each panel shows differences (continuous curve) between the two summation CDFs shown in the corresponding panel of figure 26.2. Broken curves show the standard errors of the differences, based on between-subject variation, and computed at each 10-ms interval. In each panel the vertical line segments mark the means over the two summation CDFs of five quantiles, $t_{.10}$, $t_{.25}$, $t_{.50}$, $t_{.75}$, and $t_{.90}$, to facilitate comparison across data sets.

based measures of location, spread, skewness, and kurtosis; and moments. For details of the calculation method, see the discussion of equation (21) in the Appendix.

Comparison of Proportions For proportions, we compared the values of each pair of empirical CDFs by determining their difference at each 10-ms interval. To facilitate seeing pattern in the deviations, we display the differences in figure 26.3, together with pointwise $\pm SE$ bands based on between-subject variation. There is a suggestion of a difference that is positive at low p-values and negative at high p-values, but the differences are small, and not consistent among subjects. The impression of excellent agreement is confirmed by tests: at each 10-ms time value we performed a t-test on the difference between mean proportions, using the SEs shown in figure 26.3; none of the tests was significant. Such tests can of course be performed only where the SE is nonzero, which excludes any time value at which there is a zero difference between summation CDFs for all subjects. The number of tests performed is

Table 26.4 Summation Test: Quantile Measures of Location, Scale, and Shape

	MED		IQR		SK		KR	
Experiment	*Diff* $\pm$ SE_b	*Mean*	*Diff* $\pm$ SE_b	*Mean*	*Diff* $\pm$ SE_b	*Mean*	*Diff* $\pm$ SE_b	*Mean*
1	0.2 ± 0.9	434	-0.2 ± 2.6	38	1.3 ± 7.2	30	0.9 ± 1.9	18
2 ($na = 2$)	0.9 ± 5.0	705	0.1 ± 2.9	91	3.7 ± 5.5	18	0.8 ± 0.5	16
2 ($na = 8$)	1.2 ± 3.8	894	3.2 ± 2.1	64	-5.3 ± 7.4	18	-0.4 ± 0.5	16

We estimated quantiles for $T_{11,22}$ and $T_{12,21}$, and combined them to provide robust measures of location (median, MED), scale (interquartile range, IQR), and shape (skewness, SK, and kurtosis, KR) for each distribution. For each statistic is shown the difference between the two estimates (*Diff*), a standard error SE_b of the difference based on between-subject variation, and the mean of the two estimates. See the text for definitions of SK and KR. The unit is 1 ms.

thus a conservative estimate of the number of opportunities for the SIstage model to fail. For experiments 1, 2, $na = 2$, and 2, $na = 8$, respectively, there were 42, 66, and 61 tests; they are unlikely to be independent, of course.

We also did hundreds of exploratory tests on the differences between the summation distribution proportions for individual subjects. *P*-values were not corrected for numbers of tests (Miller 1986; Johnson and Tukey 1987), so they should be taken as indications only. Among all the statistics of the CDFs we examined (proportions, quantiles, and moments), the closest we came to finding systematic deviations was in *t*-tests of the differences between proportions at each 10-ms time value for which the *SE* was nonzero. In experiment 1, none of the 173 tests was significant. In experiment 2, $na = 2$, 18 of the 318 tests were significant; in experiment 2, $na = 8$, 15 of the 241 tests were significant. These numbers would of course be consistent with the hypothesis of no difference if the tests were independent, but they are not, and the similarity of the patterns suggests that although minor, the deviations may be meaningful.[17]

Comparison of Quantiles We compared each of a set of quantiles of the two distributions, t_p for $p = .05, .10, .25, .50, .75, .90, .95$. We also compared three different functions of those quantiles, a robust measure of spread, $IQR \equiv t_{.75} - t_{.25}$, a robust measure of skewness, $SK \equiv t_{.95} + t_{.05} - 2t_{.5}$, and a robust measure of kurtosis, $KR \equiv 10(t_{.95} - t_{.75} + t_{.25} - t_{.05})/IQR$. In each case we determined differences between corresponding values for $\hat{F}^*_{11,22}$ and $\hat{F}^*_{12,21}$. Table 26.4 shows that in no case were the differences between distributions in location, spread, or shape measures significant, and in most cases the differences were small relative to the size of the quantities compared. Within-subject tests, based on between-session (experiment 1) or between-stimulus (experiment 2) variation, also failed to find significant ($p < .05$) differences between members of any of the four pairs of measures.

Comparison of Moments The interaction contrasts of means and variances, in tables 26.2 and 26.3, respectively, are equivalent to testing the differences between the first two moments of $\hat{F}^*_{11,22}$ and $\hat{F}^*_{12,21}$; we also examined the

third and fourth moments and found none of the differences to be significant, in either within-subject or between-subject tests.

26.9 THE MIXTURE TEST: FURTHER EVIDENCE AGAINST THE ALTERNATE-PATHWAYS MODEL

Based on the variance test, we noted in section 26.7 that the AP model can be decisively rejected for these data. For two reasons, however, we shall discuss application of the mixture test to the data from experiments 1 and 2: first, because the mixture test appears to be more powerful in cases where the model is wrong but where the variance test is not decisive, the mixture test would be more persuasive where the model is correct; and, second, because the pattern of deviations is instructive.

The Test Procedure

Adjustment of Distributions Variation in a factor that interacts with both A and B will not cause the mixture test to fail (section 26.3), as it could the summation test; hence we did not divide the data into subsets within subjects. For each subject we estimated the four component CDFs and averaged them in pairs to obtain $F_{\mathbf{mix}(T_{11},T_{22})}(t) = \frac{1}{2}[F_{11}(t) + F_{22}(t)]$ and $F_{\mathbf{mix}(T_{12},T_{21})}(t) = \frac{1}{2}[F_{12}(t) + F_{21}(t)]$. To average the CDFs over subjects for a graphical display free of artifacts (features in the averages but not in the individual distributions) we eliminated between-subject differences in location and scale before averaging. To do this while preserving the relation between the two mixture distributions we applied the same transformation to the two mixture distributions for each subject, a transformation defined for each subject so as to equate the mean locations and mean scales of the two distributions across subjects.[18] We then averaged the transformed CDFs to obtain those shown in figure 26.4.

Results of the Test

All three data sets show the same systematic pattern of differences between the two mixture distributions: instead of being equal, the left side of equation (9) exceeds the right for short RTs, and the right side exceeds the left for long RTs. As discussed in section 26.3, this is the pattern expected from a stage model in which the duration of each stage responds to an increase in level of the corresponding factor by lower frequencies of short durations and higher frequencies of long ones, and in which, unlike the AP model, both stages contribute to all trials.

In all three cases shown in figure 26.4, the spread appears to be greater for $F_{\mathbf{mix}(T_{11},T_{22})}$ than for $F_{\mathbf{mix}(T_{12},T_{21})}$; indeed, it can be shown that the difference between the two variances is the same as the statistic used in the variance test. However, because the mixture distributions must be identical in every respect, according to the AP model, determination of the mixture distributions provides the opportunity of using robust estimates of spread. We found, for

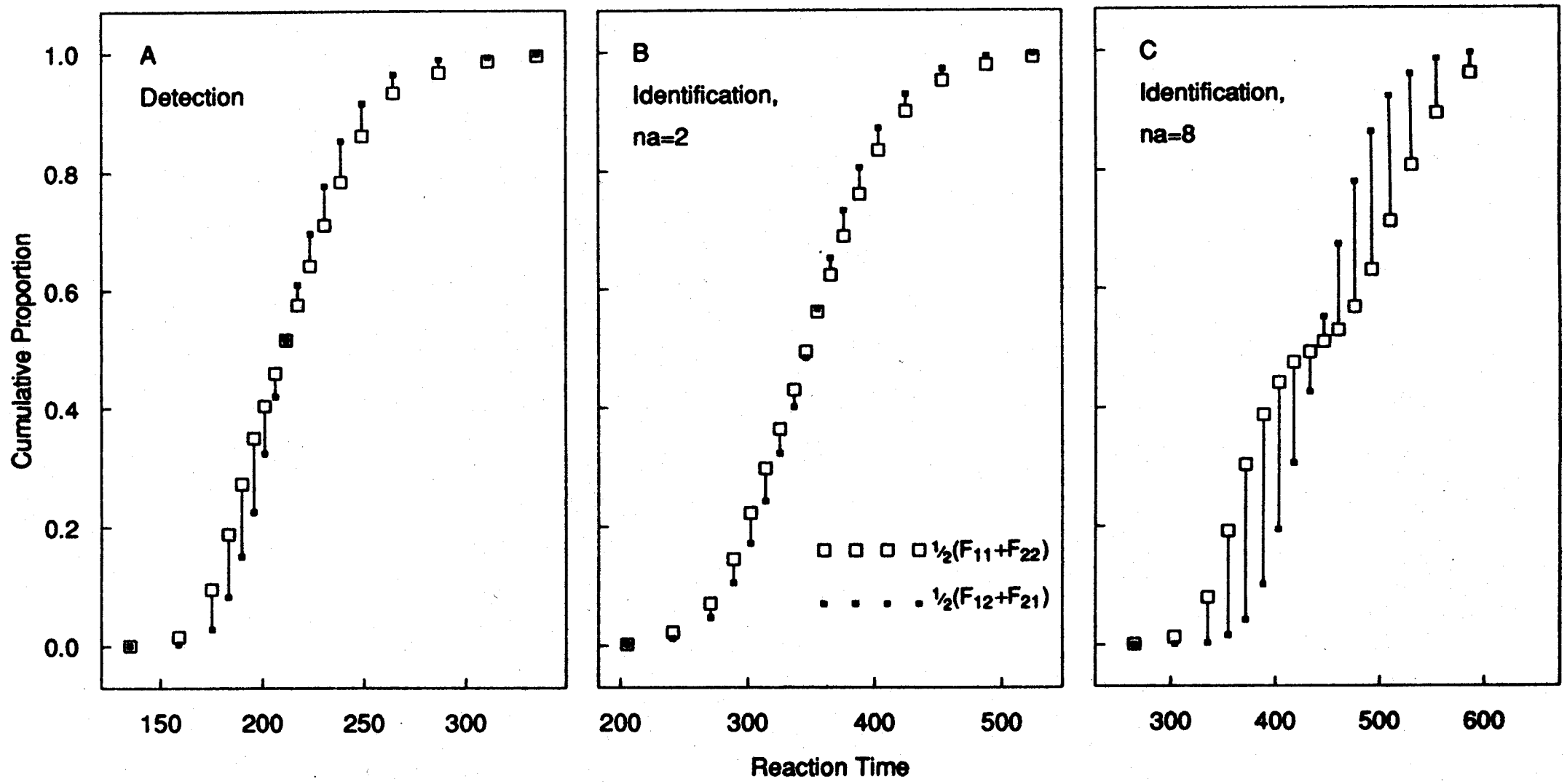

Figure 26.4 Results of the mixture test for experiment 1 (panel A) and experiment 2 (panels B, C). Mixture CDFs were adjusted and averaged as described in the text.

Table 26.5 Mixture Test: Integrated Differences between Two Mixture Distributions

Experiment	Small-RT Statistic	Large-RT Statistic	All-RT Statistic
1	$.926 \pm .054$ ($p < .0001$)	$-.854 \pm .094$ ($p = .0003$)	$.901 \pm .051$ ($p < .0001$)
2 ($na = 2$)	$.480 \pm .248$ (*ns*)	$-.793 \pm .130$ ($p = .004$)	$.757 \pm .057$ ($p = .0002$)
2 ($na = 8$)	$.996 \pm .003$ ($p < .0001$)	$-.999 \pm .001$ ($p < .0001$)	$.997 \pm .002$ ($p < .0001$)

The three statistics are defined in the text. In each case the mean $\pm$ SE_b is shown, with values of SE_b based on between-subject variation; p-values are based on two-tailed t-tests of the means versus the zeros expected from the AP model. Normalization requires all three statistics to lie within the $[-1, +1]$ interval.

example, that s^2_{bi}, a measure of spread based on the biweight (Mosteller and Tukey 1977, 208), is substantially more powerful than the sample variance, for all three data sets. We prefer, however, a test that we devised for the present purpose that would permit separate testing of conditions 17 and 18 and would be especially sensitive to the pattern of deviations from the AP model that would be expected from a stage model in which factor levels influenced the frequencies of short and long stage durations, as discussed in section 26.3.

Consider the difference between the two mixtures, $\Delta(t) \equiv F_{\mathrm{mix}(T_{11}, T_{22})}(t) - F_{\mathrm{mix}(T_{12}, T_{21})}(t)$. For a stage model in which minimum stage duration varies with factor level it is possible that $\Delta(t)$ will be positive for small RT and negative for large RT (sec. 26.3). The numerator of the *small-RT statistic* is the integrated (signed) difference $\Delta(t)$ for $0 \leq RT \leq \overline{RT}_{.5}$, where $\overline{RT}_{.5}$ is the mean of the medians of the two mixtures. If the mixtures were identical except for sampling error (AP model) then the average magnitudes of positive and negative deviations $\Delta(t)$ would be expected to be equal, so that the integrated or averaged $\Delta(t)$ taken over any interval would have zero expectation. The normalizing denominator of the small-RT statistic is the absolute difference $|\Delta(t)|$ over the same RT-range. If all the deviations were positive (negative) then the statistic would take on its maximum (minimum) value of $+1$ (-1). According to the AP model the small-RT statistic should thus be close to zero; according to the stage model and the argument associated with equation (17) it may be positive. The *large-RT statistic* is the integrated signed difference $\Delta(t)$ over the complementary interval $RT > \overline{RT}_{.5}$, similarly normalized by the integrated absolute difference over the same RT-interval. According to the AP model this also has a zero expectation; according to the stage model and the argument associated with equation (18) it may be negative. The *all-RT statistic* is the difference (small-RT statistic) — (large-RT statistic), normalized by the integrated absolute difference over the entire RT range. According to the AP model this has a zero expectation; according to the stage model it may be positive. Values of these statistics are unaffected by the distributional adjustment that we performed for purposes of graphical display.

We computed the three statistics for each subject and each of the three sets of data, with the results shown in table 26.5. Mean values of the statistics over subjects are remarkably close to their maximum possible values in several

cases. The strength of the evidence against the AP model from the mixture test is substantially greater than from the variance test (table 26.3, last column; compare significance levels); furthermore, the pattern of deviations is consistent with expectations from an SIstage model (sec. 26.3). Because they appeared to be especially sensitive, we applied the same three integrated-differences comparisons to each of the three pairs of summation CDFs (fig. 26.2). None of the nine comparisons even approached significance; the largest t value was 1.6 (with 4 df).

26.10 EVIDENCE AGAINST THE CASCADE MODEL

That several features of the SIstage model could be approximately mimicked by variants of the cascade model (sec. 26.4) led us to consider in what ways its behavior might be constrained. Instead of trying to derive such additional properties analytically, we searched for them by exploring its parameter space. We did this for each of the six models described in table 26.1. For each model, τ_c and τ_d are fixed, while τ_a and τ_b are permitted to vary. By assigning values to τ_a and τ_b we specify the model for one condition in a hypothetical experiment sufficiently well to be able to compute its distribution of cascade durations. We can then generate a random sample of the durations, or determine the expected value of any statistic. We limited ourselves to the mean and variance; for two of the models, results are illustrated in figure 26.1.

The four conditions in any 2 × 2 experiment that can be described by the model correspond to points at the corners of a rectangle in the (τ_a, τ_b) space: (τ_{a1}, τ_{b1}), (τ_{a1}, τ_{b2}), (τ_{a2}, τ_{b1}), and (τ_{a2}, τ_{b2}). For any such set of four points we can determine the expected values of four means and four variances; these eight *condition statistics* can be selected and combined in different ways to produce a single number for the experiment, an *experiment statistic*. A set of eight such condition statistics for one model is represented by the eight points given emphasis by large squares in figures 26.1B and 26.1D. Examples of experiment statistics that we have already seen are the interaction contrasts $I\{\mu_{ij}\}$ and $I\{\sigma^2_{ij}\}$.

Allowed and Forbidden Regions in a Statistic Space

For present purposes we used two new statistics, described below. With two such statistics, each hypothetical experiment is mapped onto a point in a two-dimensional *statistic space*, where each statistic is represented by a value on one dimension. In such a space the model's capabilities are represented by an *allowed region*, defined by the set of such points for all possible experiments, given the model. Results that the model cannot generate are represented by the unoccupied or *forbidden region*. Any real 2 × 2 experiment produces a single point in the two-dimensional statistic space; to test the model one asks whether this point falls into the allowed or forbidden region, taking sampling error into account.

Figure 26.1 (as well as Ashby's 1982 work) suggests that the cascade model may constrain the relation between location and spread, which is reflected in our choice of one of the experiment statistics. (Conditions that slow the model's response tend to do so by slowing the rate at which activation grows, and a slower growth rate increases the RT spread induced by the fixed criterion variance.)

Examination of the relation between location and spread requires some care in the choice of measures. First, because the time unit in the model is arbitrary, we must either estimate it or use measures that don't depend on it. Second, we must allow for the possibility that in generating the observed RTs the postulated cascade mechanism, influenced by factors A and B, is concatenated with other processes not influenced by those factors, such as the response mechanism assumed by McClelland (1979) and Ashby (1982) or supplementary stages or cascade units, processes that might contribute, along with the cascade mechanism, to both location and spread. Our measures should reflect the cascade mechanism alone, whether or not such supplementary processes are present.

To avoid distortion due to such supplementary processes we use statistics defined as factor effects, i.e., as differences between RT measures at different factor levels, exploiting the assumption that the supplementary processes are unaffected by factors A and B. Not all measures will serve, however. Thus, if RT_1 and RT_2 are RTs obtained in conditions 1 and 2, M is a measure of location or spread, and T_{C1} and T_{C2} are the contributions to the RT from the cascade mechanism, then we require that $M(RT_2) - M(RT_1) = M(T_{C2}) - M(T_{C1})$. If T_S represents the duration of the supplementary processes, this condition in turn requires that any measure be additive, in the sense that $M(RT) = M(T_C) + M(T_S)$. The mean satisfies this requirement for the location measure, but an additional assumption is needed to select an appropriate measure of spread. One possibility is the variance; another that has been applied in testing the cascade model (Ashby 1982) is the standard deviation. The variance is additive if $\mathrm{cov}(T_C, T_S)$ is zero or constant, or if the supplementary processes consist of one or more additional cascade units. Additivity of the standard deviation requires that $|\mathrm{cov}(T_C, T_S)| = \sigma_{T_C}\sigma_{T_S}$, its maximum possible value. We believe that variance additivity is more likely, so we chose the variance as our measure of variability. (Our conclusions may depend on the validity of this assumption.) Let η_{ij} and θ^2_{ij} be the mean and variance of the cascade duration under factor levels i and j, μ_S and σ^2_S be the mean and variance of T_S, and t_u be the scale factor expressing the model's time unit in ms. We can then write $\mu_{ij} = t_u\eta_{ij} + \mu_S$ and $\sigma^2_{ij} = t^2_u\theta^2_{ij} + \sigma^2_S$. To link the cascade process under study to the data, we constructed statistics that eliminated μ_S and σ^2_S (by taking differences of quantities assumed to be additive), and eliminated t_u (by forming dimensionless ratios of these differences). The first statistic relates the change in variance to the corresponding change in mean, as both factor levels are increased:

$$\text{variance-change statistic} \equiv \frac{\sigma_{22}^2 - \sigma_{11}^2}{(\mu_{22} - \mu_{11})^2} = \frac{\theta_{22}^2 - \theta_{11}^2}{(\eta_{22} - \eta_{11})^2}. \tag{22}$$

Our explorations suggest, for example, that the values of the variance-change statistic that the cascade model can produce are bounded below. The second statistic is the difference between the two main effects, normalized by their sum:

$$\text{main-effect difference statistic} \equiv \frac{|\mu_{d\cdot} - \mu_{\cdot d}|}{\mu_{d\cdot} + \mu_{\cdot d}} = \frac{|\eta_{d\cdot} - \eta_{\cdot d}|}{\eta_{d\cdot} + \eta_{\cdot d}} = \frac{|\eta_{21} - \eta_{12}|}{\eta_{22} - \eta_{11}}, \tag{23}$$

which can range from zero to one. We shall see that larger main-effect differences are associated with more severe bounds on values that the model can produce for the variance-change statistic.

The Test Procedure

We studied each of the six models described in table 26.1. For each model we determined the mean and variance of the processing time for each of the 100 points in the two-parameter grid used to create figure 26.1. Within this grid there are 2025 possible simulated 2 × 2 experiments in which both factors have nonzero effects; for each of these we computed the variance-change statistic and the main-effect difference statistic. Results of these computations for model 3.1 are shown as an example in figure 26.5A; each plotted point represents the outcome of one simulated experiment, and the allowed region of the model is included within the surrounding contour, the convex hull of the set of points. The figure, which shows the left-hand portion of this region, reveals that values of the variance-change statistic that can be achieved by the model are bounded below, and that the bound varies with the main-effect difference. The allowed regions are similar for the six models, as shown by figure 26.5B, where the contours for three of them are displayed on an expanded abscissa. The similarity suggests that this feature of the cascade model depends little on values of the fixed time constants, and is thus quite general.

We took this analysis further in three ways. First, we determined whether the quantization of the parameter grid contributed to the constraints, by sampling parameter values in .01-unit steps in the range (.05, 1.00). This caused the allowed region (dotted in fig. 26.5B) to expand upwards, but not to the left. Second, we extended the parameter range in model 3.2 to encompass a 100:1 ratio, despite its implausibility as discussed in section 26.4, again using a sampling method. The result is the leftmost contour in figure 26.5B, which reveals only a mild relaxation of the lower bound. Finally we examined the small-sample properties of the main-effect difference and variance-change statistics by using model 3.2 with several pairs of (τ_a, τ_b) values to generate random samples of the same size as in our experiments, to check for bias; we found no evidence of bias.

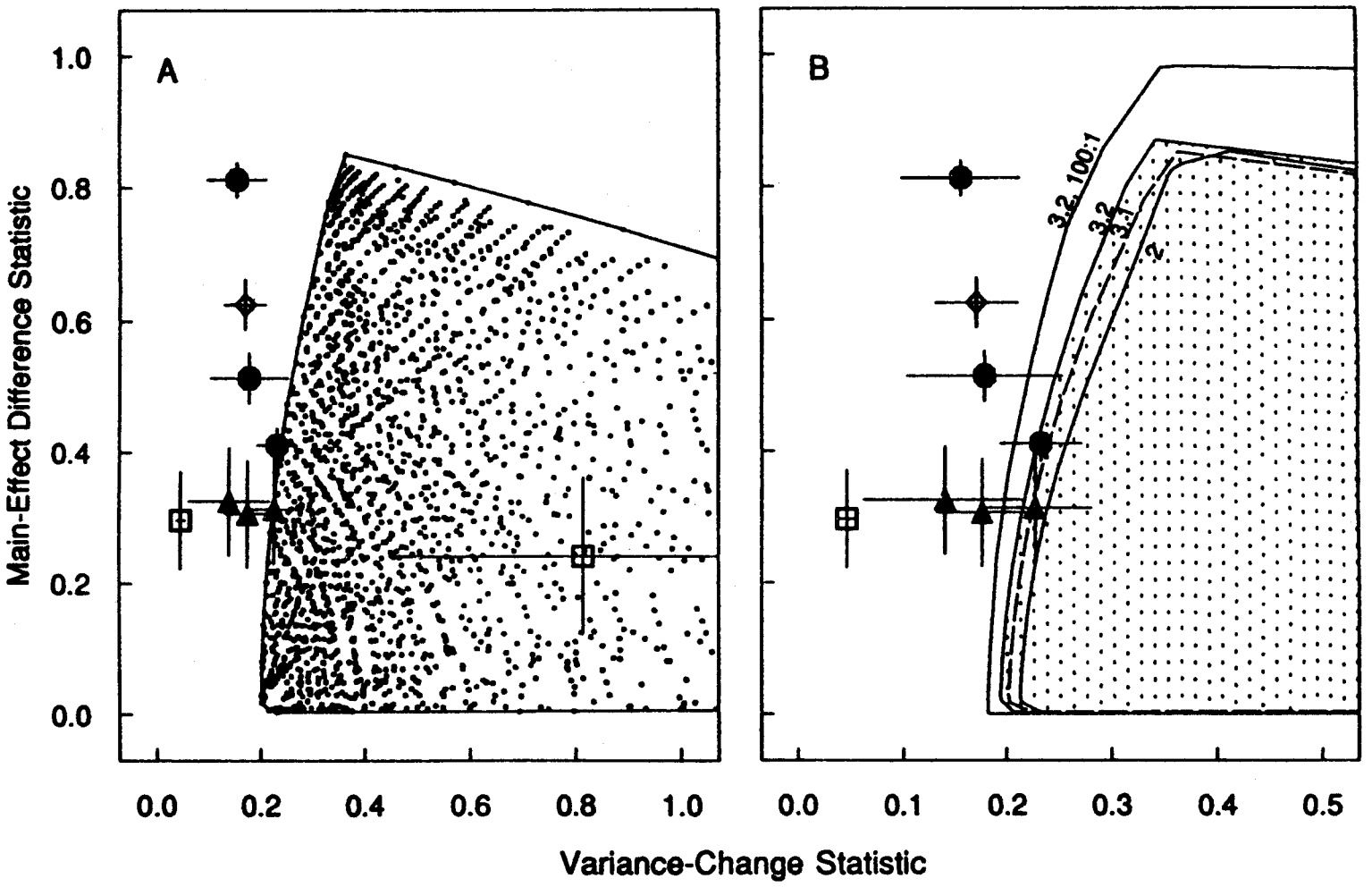

Figure 26.5 Data points versus allowed and forbidden regions of the cascade model in a two-dimensional statistic space. *Panel A:* Each of the numerous small points represents the expected values of the two statistics for one simulated experiment generated by model 3.1 (See table 26.1.) The set of 2025 points continues off the plot to the right. Surrounding the points is the contour defined by their convex hull, which includes the allowed region of the model, subject to considerations discussed in the text. Also shown by larger points are data from the four real experiments, together with lines marking $\pm SE_b$ for each statistic, based on between-subject variation. For each of experiments 1 and 3 (2 × 3 designs), there are three points, one for the corners of the design (also discussed elsewhere in this chapter), and one each for the two other possible 2 × 2 experiments. Experiment 1: filled octagons, the top point for levels of factor B (foreperiod) of 950 and 1150 ms, the middle point for 750 and 950 ms, and the bottom point for the corners, 750 and 1150 ms. Experiment 2: unfilled squares, $na = 8$ to the left, $na = 2$ to the right. Experiment 3: filled triangles, the left point for levels of factor B (set size, s) of $s = 2$ and $s = 4$, the right point for $s = 1$ and $s = 2$, and the middle point for the corners, $s = 1$ and $s = 4$. Experiment 4: unfilled diamond. *Panel B:* Note the expanded abscissa scale. Contours abutting and within the dotted area mark the allowed regions for models 2, 3.1, and 3.2, all with time-constant ratios no greater than 20:1. The contours for models 4.1, 4.2, and 4.3, with the same 20:1 constraint, fall almost entirely between those for models 3.1 and 3.2. The contour labeled "3.2, 100:1" shows the allowed region when time-constant ratios are permitted to increase to 100:1. Data points as in panel A; experiment 2, $na = 2$, is now off scale.

Test Results: Data versus Allowed Region

Also shown in figures 26.5A and 26.5B are points that represent the same pair of statistics for each of the four experiments, together with lines indicating $\pm SE$ based on between-subject variability. For the detection and classification experiments, in which one of the factors had three levels, the figure shows the statistics for each of the three possible 2 × 2 designs. (Within each of these two sets of three, values are not independent.) For most of the data points the variance-change statistic is too small for the model.

Excess Model Variance in the Best-Fitting Case

To gain further insight into the cascade model we considered the version of model 4.3 ($\tau_a = 0.04, 0.4$; $\tau_b = 0.05, 0.95$) that provided the best fit within the two-dimensional statistic space (fig. 26.5) to the data from experiment 1 with levels of factor A of 750 and 1150 ms. When we adjusted the time unit t_u in the model by forcing it to produce agreement with the separation between the means of conditions 11 and 22, we found that the average variance of the distributions produced by the model is about twice as great as the variance in the data. Thus the variances in the model are too large relative to the main effects on the means. This can be seen by comparing the simulated distributions on the left side of figure 26.6 with the actual distributions in figure 26.2A. Both the upper and lower tails of the four distributions tend to be more separated in the data than in the model, where they appear to fan out from common points. This property may reflect an essential feature of the cascade model: its continuous transmission of activation. Because activation in the *n*th unit starts rising immediately when the stimulus is presented, the increment in activation that triggers the response occasionally occurs shortly after the stimulus is applied, even in the slowest condition. And because activation in the fixed slow process needed for additivity continues to rise for a long time, the model produces a few slow responses, even in the fastest condition. Adjustments of the model that might reduce the discrepancy include reducing the spread of the criterion distribution, or changing its shape so as to introduce a high threshold for triggering the response.

Application of the Summation Test to a Cascade Mechanism

On the right side of figure 26.6 is shown the result of applying the summation test to the four simulated distributions.[19] Consistent with Ashby's (1982) findings, the test works remarkably well. Thus, just as mean and variance additivity fail to discriminate cascade and stage models, so does the summation test (though this is known only under limited conditions); this observation emphasizes the importance of discovering necessary conditions for that test.

One possibility is that the conditions that permit the cascade model to emulate a stage model (the incorporation of one or more fixed units with long time constants) are also ones under which it approximates a stage model in

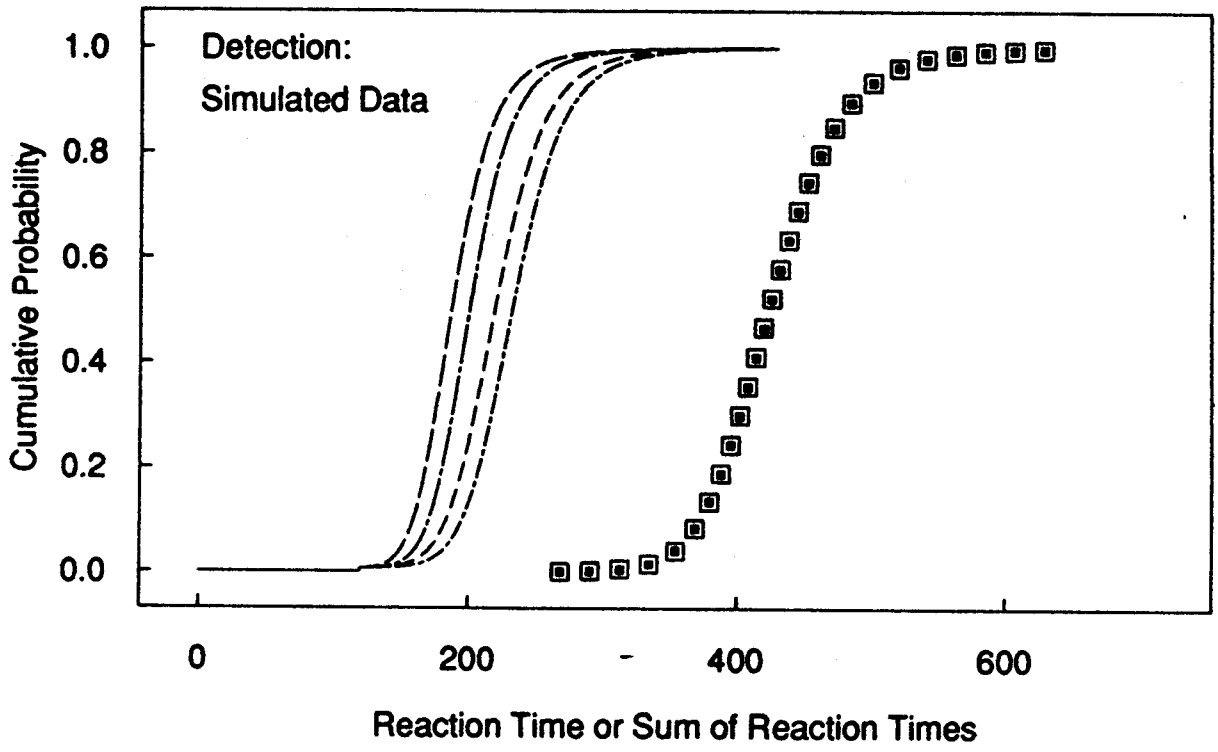

Figure 26.6 Summation test applied to the cascade model. Results of simulation by cascade model 4.3 of the data from experiment 1 (detection) shown in figure 26.2A. Simulated RT distributions from the four conditions are shown on the left, with a constant (121 ms) added to all simulated times; the two convolutions of these distributions, obtained by applying numerical convolution to the pairs of density functions $\{f_{T_{11}}, f_{T_{22}}\}$ and $\{f_{T_{12}}, f_{T_{21}}\}$ are shown on the right. The maximum (vertical) separation between the two functions on the right is .001.

function. In the defining conditions of a stage model, process *a* has completed its work before process *b* starts, and, assuming selective influence, the time required by process *b* is independent of systematic changes in the rate of process *a*. If we think of the slow unit in a cascade model as placed between units *a* and *b*, then it seems possible that its presence causes these two stage-model conditions to be approximately satisfied. This conjecture is encouraged by some exploratory comparisons of the activation functions of unit *a* alone, and of units *a*, *b*, and *c* together, using several different proportions of asymptotic level to define a process having "completed" its work; the two conditions were approximately satisfied.

Future Tests

Quantitative analyses of the cascade model, such as those above, may help to guide future tests. For example, reference to figure 26.1 will suggest why experiments with more than two levels per factor may challenge the cascade model more severely: Suppose that an increase in level of factor A induces a given change in mean RT by changing τ_a from τ_{a2} to τ_{a3}. Because the function relating variance to τ accelerates, the corresponding change in variance is greater if τ_{a2} is greater. Requiring the model to fit a third, faster condition (yielding τ_{a1}) as well, will force τ_{a2} and τ_{a3} to increase, which in turn will increase the variance-change statistic associated with levels 2 and 3, and hence perhaps the disparity between model and data. Future tests using 2 × 2 designs can also be guided by figure 26.5, which shows that the model

is challenged more by experiments with unequal main effects. (Shaw 1984 provides another example of the use of quantitative theory to guide experimental design.)

26.11 DISCUSSION

Additivity of the effects of experimental factors on mean RT has often been taken to suggest the existence of stages. In this chapter we have considered two alternative explanations of additivity, one new (the alternate-pathways model) and one developed earlier (the cascade model). That additivity has more than one possible explanation has long been taken for granted, at least by critics of the additive-factor method, and is hardly surprising. Nevertheless, the AP model is the first alternative to the stage model we know of that produces exact additivity; the cascade model approximates additivity with some parameter values but not others. The AP model has the virtue of being simple and plausible, yet readily testable because it strongly constrains the data. It is noteworthy that three models so different in spirit can explain the same pattern of means. Because of this, discriminating among the models has required analysis of other aspects of the data. Thus we have tried to discover what the models predict about variances and entire distributions, and have confronted these predictions with five sets of data from a wide range of RT experiments.

The most important feature of this work is the new evidence for stage models. This consists of: the dramatic failures of the AP model for all five data sets (tables 26.3, 26.5; fig. 26.4); failures of the cascade model (figs. 26.5, 26.6); and successes of the SIstage model (tables 26.3, 26.4; figs. 26.2, 26.3, 26.4). The successes of the summation test (table 26.4; figs. 26.2, 26.3) seem especially persuasive.[20] It should be kept in mind that the summation test was applied to only two experiments (detection and identification), which, moreover, have features in common, such as verbal instructions and visual stimuli. However, the two experiments also differ in several ways, such as the meaningfulness of the stimuli, the number of possible responses, the nature of the responses (the muscles used, for example), and the ranges of RTs.

One important limitation of the present work is that it is not altogether clear what success of the summation test implies: the necessary conditions—stronger than the Gstage model, but weaker than the SIstage model—have yet to be discovered. We do not know why the cascade model satisfies the test under some parameter settings; perhaps it is because with those parameters it resembles a stage model. Another important limitation of this work lies in our uncertainty about how to interpret the failures of the cascade model. They may reflect fundamental constraints of an interesting class of models of which it is a member, or may depend on details, such as the criterion distribution or the law that governs the growth of activation. Rejection of the cascade model for these data may thus be less interesting than rejection of the much simpler, and in that sense, weaker AP model. One interesting issue here is how the failure of the AP model bears on the validity of the complementary set of

mixture models (Yantis, Meyer, and Smith 1991) in which the pathway probabilities vary while the pathways remain fixed. A final limitation worthy of mention is that the "off the shelf" experiments that provided the data we used had not been designed to discriminate among the three models. It seems likely that future work will suggest better experiments for this purpose; factors with more than two levels, and experiments with more than two factors, for example, are likely to be helpful.[21]

When alternatives to the stage model have been proposed in the past, it has not been clear how they constrain the data, that is, what they predict. Nor has it been clear what the stages explanation of additivity predicts, although the work of Meyer et al. (1984, 1988), Miller (1988), and Schweickert (1985), for example, has taken important steps in that direction. If an explanation is based only on the result that suggested it, it may not be especially convincing, and a preference for that explanation over competitors may be only a matter of taste. For example, Taylor (1976) proposed that additivity may arise because the factors changed one stage "in an additive way" (181). Because it resembles the fact it explains, and predicts nothing, this is not a satisfying explanation. However, as long as competing explanations also predict nothing, it is hard to dismiss. It is much easier to dismiss when a competing explanation makes a successful prediction. In particular, the summation test, based on Ashby and Townsend's (1980) suggestion, will help compare stage models to competitors. A traditional difficulty associated with testing the properties of RT distributions other than their locations is the large sampling error associated with the variance and higher moments, especially given the high tails and outliers of some distributions. Along with the mixture test developed for the AP model, the summation test permits the use of robust measures of spread, skewness, and so forth, which are likely to be less variable than the sample moments.

The evidence for the existence of something like stages (independently changeable, serially arranged operations) is now both deep and broad. It is deep—relatively extensive and persuasive—for the detection and identification experiments. It is broad—relatively general—because Roberts (1987) found evidence for stages in animal response-rate experiments, which have little in common with human RT experiments beyond the use of vertebrate subjects.

The improvements in the evidence have come at roughly the same time that theorizing in human experimental psychology has shifted toward models of high complexity. Astrophysicists have a joke: Never propose a theory that can be tested in your lifetime. By proposing complicated theories with many free parameters, psychologists seem to be moving in all seriousness quite close to this ideal. Complex models are usually "tested" by asking if they can fit another data set. The virtue of such models, in the eyes of their proponents, is often the range of the data they can fit. This is indeed a virtue; a model that explains only one result is not very interesting. Such tests are incomplete, however. They should increase our belief in the model only if they are sufficiently extensive to identify forbidden as well as allowed regions in the

statistic space. This requires knowing what the model cannot do, as well as what it can (cf. Massaro 1988). And determination of what a complex model cannot do may require a vast amount of computation.

The shift toward relatively complex models has been motivated, at least in part, by considerations of plausibility. When choosing a theory, this is a good place to start. The message of this chapter, however, is that something implausible is apparently true: The vertebrate brain, in a wide range of situations, has a simple functional structure.

APPENDIX: PROOFS AND COMMENTS ON EQUATIONS

Equation (3): Additivity of factor effects follows from additivity of stage durations and its preservation under expectation, and from the selective influence of factors A and B on T_{ai} and T_{bj}, respectively. Take expectations in equation (1), replace the μ_{ij} in equation (3), and collect terms. (See Sternberg 1969.)

Equation (5): In a proof by construction, which also suggests a method of testing, note that $T_{11} + T_{22}$ and $T_{12} + T_{21}$ can each be regarded as the sum of RTs that results from concatenating two stage models; the two sums are the same because the two concatenations contain the same stages at the same factor levels, just differently ordered. To elaborate, $T_{11} + T_{22} = T_{a1}[+]T_{b1} + T_{a2}[+]T_{b2}$, and $T_{12} + T_{21} = T_{a1}[+]T_{b2} + T_{a2}[+]T_{b1}$. Here, "+" represents summation of RTs from different conditions, which are stochastically independent by construction (sec. 26.8). In the first sum, for example, the construction method implies that T_{a1} and T_{b1} are each independent of T_{a2} and T_{b2}. Summation of stage durations that contribute to the same RT is represented by "[+]." Given stochastic independence of stage durations, the two kinds of summation are equivalent. Because the two sums are composed of the same independent random variables, they have the same distribution.

In a more formal proof, we note that summation of random variables (RVs) requires them to be defined over the same sample space. For the original RVs, $\{T_{ai}\}$ and $\{T_{bj}\}$, there are, instead, four different sample spaces, one each for conditions 11, 12, 21, and 22. For example, T_{a1} corresponds to two RVs, one in condition 11 and one in condition 12; let us call them T_{a11} and T_{a12}, respectively. By the assumption of selective influence, factor B has no effect on T_{a1}; hence T_{a11} and T_{a12} are identically distributed. To permit the desired summation we first replace each of these identically distributed pairs of original RVs by a new RV with the same distribution, such that the four new RVs are mutually independent and are defined over the same sample space. For example, we replace T_{a11} and T_{a12} by T'_{a1}. (It is well known that such new RVs whose distributions are the same as the marginal distributions of the original RVs can be defined.) Because of the assumed stochastic independence of the original RVs (such as T_{a11} and T_{b11}, which sum to T_{11}), the sum of each pair of new RVs (such as $T'_{a1} + T'_{b1}$) has the same distribution as the sum of the corresponding pair of original RVs. Now, however, T_{11} and T_{22}, for example, are defined over the same sample space, thus permitting us to define

their sum, which we write as above, except that the $\{T_{ai}\}$ and $\{T_{bj}\}$ are replaced by the $\{T'_{ai}\}$ and $\{T'_{bj}\}$.

Given the independence of the $\{T_{ij}\}$, equation (5) is equivalent to additivity of cumulants of all orders: $I\{\kappa_{rij}\} = 0$, $r \geq 1$ (Sternberg 1969). However, because of instability of their estimates, testing equation (5) by using the higher cumulants is less attractive than using other aspects of the distributions, which the summation test makes possible.

Equation (6): To show that equation (5) is implied by this translation condition plus means additivity, observe that the condition implies $T_{11} + T_{22} = \mu_{11} + \mu_{22} + t + t$ and $T_{12} + T_{21} = \mu_{12} + \mu_{21} + t + t$, where the subscripts on the $\{t_{ij}\}$ are dropped to emphasize the equality of their distributions. Given means additivity, $\mu_{11} + \mu_{22} = \mu_{12} + \mu_{21}$, so the two sums have the same distribution.

Equation (8): Use equation (7) to expand each term, and collect terms. As for equation (5) of the SIstage model, there is a proof by construction that also suggests a method of testing. Write equation (8) as equation (9), to which it is equivalent. The left side of this equation is the CDF of an equal-probability mixture of the processes that generate T_{11} and T_{22}, and the right side similarly for T_{12} and T_{21}. The two corresponding mechanisms each have four alternate pathways with pathway probabilities $\frac{1}{2}p$, $\frac{1}{2}(1-p)$, $\frac{1}{2}p$, and $\frac{1}{2}(1-p)$; the two mechanisms are equivalent because they differ only in the ordering of the same four pathways. The suggested *mixture test* is to pool equal-size samples from T_{11} and T_{22}, and from T_{12} and T_{21}; CDFs of the resulting data sets should be equal, except for sampling error. If sample sizes are unequal, then comparison of the means of pairs of corresponding CDFs permits better use of the information in the data. As in the summation test, equality of CDFs implies identity of any statistic, so that one can choose measures with desirable sampling properties, for example.

Equation (11): Differentiate equation (8) with respect to t to get density functions, multiply by t^r, and integrate with respect to t.

Equation (14): Additivity of the means ($I\{\mu_{ij}\} = 0$) permits us to replace μ_{12}, μ_{21}, and μ_{22} in $I\{\mu_{ij}^2\}$ by $\mu_{11} + \mu_{1d}$, $\mu_{11} + \mu_{d1}$, and $\mu_{11} + \mu_{1d} + \mu_{d1}$, respectively. Do so, expand, and collect terms.

Equation (15): This failure of variance additivity for the four component distributions is equivalent to equality of variances of the two mixture distributions specified in equation (9), the latter, however, can be tested by using measures of spread that may have more desirable sampling properties than the variance.

Equation (16): The variance of an equal-probability mixture of X and Y is $(\sigma_X^2 + \sigma_Y^2)/2 + (\mu_X - \mu_Y)^2/4$. Use this to rewrite the terms on the left, and use variance additivity to eliminate the terms in σ^2. Now expand, and use equation (14) and its method of proof.

Equation (17): From equation (8), $F_{12}(t) + F_{21}(t) = F_{11}(t) + F_{22}(t)$. Now note that $F_{22}(t) \geq 0$, $(t \geq 0)$.

Equation (18): From equation (8), $\bar{F}_{12}(t) + \bar{F}_{21}(t) = \bar{F}_{11}(t) + \bar{F}_{22}(t)$. Now note that $\bar{F}_{11}(t) \geq 0$, $(t \geq 0)$.

Equation (19): See McClelland (1979).

Equation (20): See Ashby (1982) for proof of an equivalent result. To generalize from a criterion distribution with unit variance to one with variance σ_c^2, replace $\Phi(x)$ by $\Phi(x/\sigma_c)$ in both numerator and denominator. Because the normal criterion distribution has nonzero probability when $t < 0$, $G_n(t)$ has accumulated probability .006 at $t = 0$. In our computations we have set $G_n(0) = 0$. To compute the expected mean and variance we used a discrete approximation of equation (20), truncated at a t-value great enough for sufficient accuracy, then differentiated numerically to obtain the density function, then summed after multiplying by t^r for the rth raw moment. For a random sample from the $G_n(t)$ distribution we used $G_n^{-1}(U)$, where U is the value of a random variable distributed uniformly on the [0, 1] interval.

Equation (21): Calculations for the summation test, including this adjustment, were performed in the following order: (a) the data were divided into subsets within each subject, each subset containing RTs from four conditions, and indexed by k as described in the text. Then, for each subset, (b1) the Cartesian product method (see text) was used to provide two summation sets; (b2) the median and interquartile range were computed for each summation set, and averaged over the two summation sets, giving λ_k and ξ_k; (b3) λ_k and ξ_k were used in equation (21) to adjust each value in the two summation sets, providing two *adjusted summation sets*; (b4) the statistic—proportion, quantile-based measure, or moment—was computed for each of the two adjusted summation sets; (b5) the difference between the two values of the statistic (one for each summation set) was found. Finally, (c1) for each subject the differences were combined over subsets to provide a mean and a within-subject SE; (c2) for the set of subjects, the subject means were combined to provide a final mean and a between-subject SE.

Equations (22) and (23): Replace the μ_{ij} and σ_{ij}^2 by their equivalents in η_{ij}, θ_{ij}^2, μ_S, σ_S^2, and t_u, and combine terms.

NOTES

Supported in part by Grant AFOSR-91-0015 from the Air Force Office of Scientific Research to the University of Pennsylvania, and by a Sloan Foundation Research Fellowship to Seth Roberts. We thank Sylvan Kornblum and David E. Meyer for encouragement and helpful comments. We are grateful to Paul Rosenbaum and Frank Norman for statistical and mathematical guidance, to Benjamin T. Backus, James C. Johnston, and Robert S. McCann for their generosity in permitting us to use unpublished details of their data, and to F. Gregory Ashby, David H. Krantz, Richard Schweickert, James T. Townsend, and Dirk Vorberg for comments on an earlier draft.

1. The theory below is limited to situations in which there are just two factors and two underlying processes, and where each factor has just two levels. It is straightforward to extend these developments to cover experiments with more than two factors, and/or more than two levels per factor. This is done to some extent in the data analyses that follow.

2. Given two or more mechanisms, each of which produces additive effects of the same pair of factors on mean RT, hybrid mechanisms can be created for which that property is preserved. One type of hybrid is the serial concatenation of two such mechanisms; another is a probability mixture. Consideration of such hybrids is an interesting problem for future work.

3. Miller, van der Ham, and Sanders (unpublished manuscript) have drawn the same conclusion.

4. See the Appendix for proofs and comments on equations.

5. The summation test is a stronger test of the stage model than merely the additivity of means and variances. That is, equations (3) and (4) are not sufficient for equation (5). Again we thank Frank Norman for an example: Suppose that both of two factors have nonzero effects, and that F_{11}, F_{12}, F_{21}, and F_{22} satisfy equations (3) and (4). Then if the F_{ij} are all Gaussian distributions, equation (5) is satisfied. But if the F_{ij} are all gamma distributions, equation (5) also requires that the differences among them are limited to their shape parameters; their scale parameters cannot differ.

6. Alternative pathways should be distinguished from parallel processes. However, an AP mechanism could result from a pair of self-terminating parallel processes if there were fluctuating dominance of one process over the other, with no overlap between completion times of the dominant and nondominant process, and if the probability of dominance was unaffected by factors A and B. Alternating cerebral dominance (Weintz et al. 1983) exemplifies a mechanism that might produce such fluctuating dominance of processes.

7. Unless stated otherwise, the u in $u \pm v$ represents the mean over subjects, and the v represents SE_b, the standard error based on between-subject variation.

8. The present discussion is concerned with population minima and maxima and is intended for heuristic purposes; we do not consider tests based on sample extrema.

9. McClelland (1979) assumed that the noise in the final output activation level was distributed as a standard normal distribution ($\mu = 0, \sigma^2 = 1$). Such noise can alternatively be associated with the criterion, as in Grice's (1972) account of response evocation (see also Grice, Canham, and Boroughs 1984; and Luce 1986, section 4.3); the criterion is then normally distributed with unit variance $\sigma_c^2 = 1$ about its mean. The important constraints in either interpretation are that the variance of the noise distribution is independent of the time constants, and that only one sample from the distribution is used per trial. The time added by the response mechanism is assumed to be stochastically independent of the other component of RT, and is not included in the calculations that lead to equations (19) and (20).

10. The choice of values was constrained by the cascade equation's (19) requirement that all rate parameters be distinct.

11. If the slow unit is influenced by factor C, then a test is provided by lowering the level of C, which may convert an additive relation between A and B into an overadditive one.

12. Ashby (1982, table 2) also found indications of good means additivity, but parameter values in most of his examples caused one of the main effects to be substantially smaller than the other, which tends to be associated with small interactions, and interactions were not measured relative to main effect sizes. Ashby also reported that under some conditions the cascade model could approximately satisfy the summation test. As discussed in section 26.3, this implies failure of the mixture test.

13. For $\mu_1 < \mu_2 < \mu_3$ the model requires that $(\mu_3 - \mu_1)/(\mu_2 - \mu_1) < (\sigma_3^2 - \sigma_1^2)/(\sigma_2^2 - \sigma_1^2)$.

14. Unless stated otherwise the significance level is $p = .05$.

15. The Cartesian product of two sets consists of all possible pairs containing one member of each set. If one set has m members, and the other, n, the Cartesian product consists of $m \times n$ pairs.

16. The Kolmogorov-Smirnov (K-S) test of distributional equality would be applicable if the summation set for $T_{11,22}$, for example, contained just the 31 statistically independent values obtained from one of the 31^2 possible random pairings of the observations in conditions 11 and 22. The Cartesian-product summation set uses more of the information in the data, but the resulting 961 values do not meet the K-S test's requirement of statistical independence. The

summation CDFs can be regarded as smoothed versions of the estimates based on single random pairings, for which sampling properties of the K-S test statistic are likely to differ. As Grayson (1983) observed, the same concern about applicability of the K-S test applies to empirical CDFs generated by density estimation and numerical convolution (Ashby and Townsend 1980). At this writing, unfortunately, there appears to be no appropriate replacement test. A test of the identity of the two Cartesian-product CDFs is needed that is especially sensitive to likely departures, such as the differences in spread expected from the AP model, rather than broad-gauge and sensitive to location differences, as is the K-S test.

17. The deviations were shown by the same two of the five subjects in the $na = 2$ and $na = 8$ conditions. In each case the significant tests were located at a set of contiguous time values (separated by 10 ms). They were of the same sign within subjects, but different between subjects. For one subject the ten (six) significant tests for $na = 2$ ($na = 8$) were associated with mean summation CDF values from about .02 to .29 (.12 to .47); the median significance level was .003 (.03); and the differences $\hat{F}^*_{11,22}(t,k) - \hat{F}^*_{12,21}(t,k)$ were negative, with a mean of $-.04$ ($-.08$). For the second subject the eight (nine) significant tests for $na = 2$ ($na = 8$) were associated with mean summation CDF values from about .04 to .25 (.01 to .38); the median significance level was .03 (.004); and the differences were positive, with a mean of .06 (.06).

18. We first estimated the median and interquartile range of each of the mixture CDFs for each subject. Let λ_k and ξ_k be the means of these medians and IQRs, respectively, for subject k; let $\lambda_.$ and $\xi_.$ be their means over subjects; and define $t^*_k \equiv \lambda_k + (\xi_k/\xi_.)(t - \lambda_.)$, a linear transformation of t. The pair of transformed mixture distributions for each subject, $F^*_{\mathrm{mix}(T_{11},T_{22})}(t)$ and $F^*_{\mathrm{mix}(T_{12},T_{21})}(t)$, produced by applying the same transformation $F^*_k(t) \equiv F_k(t^*_k)$, then has the desired property of having $\lambda_.$ as their mean median and $\xi_.$ as their mean IQR for all subjects.

19. For the method (convolution by fft), see Press, Flannery, Teukolsky, and Vetterling 1988, section 12.4.

20. Note added in proof: To strengthen the summation test we have extended the search for differences between $F_{T_{11}+T_{22}}(t)$ and $F_{T_{12}+T_{21}}(t)$ to histograms and density estimates, to supplement what we had already done for CDFs, and we have fitted orthogonal polynomials to the CDF differences to render the tests independent. These new tests strengthen our conclusions. We have also begun to explore the sensitivity of the summation test to violations of the SIstage model by applying it to data that simulate the AP model, thereby introducing stochastic dependence of stage durations. In an attempt to be realistic we created such data from experiment 1, experiment 2, $na = 2$, and experiment 2, $na = 8$, by retaining the T_{11} and T_{22} values in each data set, but replacing T_{12} and T_{21} by random halves of the values obtained by pooling T_{11} and T_{22}. The summation test failed dramatically for all three simulations. For example, in comparisons of proportions (analogous to those in fig. 26.3, where there were no significant differences among 169 tests) there were 39, 69, and 91 tests for the three data sets, of which 29, 37, and 57 were significant, respectively. In the IQR comparisons (analogous to those in table 26.4) the differences and means were (24.0 ± 2.6, 51.8), (16.7 ± 5.0, 97.9), and (102.4 ± 1.9, 116.9), with corresponding significance levels $p = .0002$, $p = .03$, and $p < .0001$, respectively. These results show that our procedures are sensitive to at least one type of violation of the SIstage model of a size that might be observed in actual experiments.

21. Consider a three-factor experiment with each factor at two levels and assume that the effects of the factors on mean RT are pairwise additive. The eight conditions can be represented as vertices of a cube, where the three dimensions (x, y, z) are levels of the three factors. Each of the six faces of the cube provides one two-dimensional summation test. For example, the front face, where $z = 1$, provides $T_{111} + T_{221}$ and $T_{121} + T_{211}$ (which should have the same distribution). Similarly the right face, where $x = 2$, provides $T_{211} + T_{222}$ and $T_{212} + T_{221}$. These are just particular cases of the test we have already used. In addition, there are six distributional equality relations among four sums of RTs for conditions represented as vertices of oblique planes within the cube, such as $T_{111} + T_{222}$ and $T_{112} + T_{221}$. The proof that these two sums (and two other similar pairs, $T_{121} + T_{212}$ and $T_{122} + T_{211}$) all have the same distribution, given the SIstage

model, is similar to the proof of equation (5). These twelve summation tests reflect a basis containing four orthogonal tests. One such set of four are those corresponding to the left, front, right, and bottom faces of the cube. Thus the extra additive factor provides a disproportionate increase in the number of possible summation tests of the SIstage model: One $2 \times 2 \times 2$ experiment provides twice as many independent tests as two 2×2 experiments.

REFERENCES

Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review, 89,* 599–607.

Ashby, F. G., and Townsend. J. T. (1980). Decomposing the reaction time distribution: Pure insertion and selective influence revisited. *Journal of Mathematical Psychology, 21,* 93–123.

Atkinson, R. C., and Juola, J. F. (1974). Search and decision processes in recognition memory. In Krantz, D. H., Atkinson, R. C., Luce, R. D., and Suppes, P. (Eds.), *Contemporary developments in mathematical psychology.* Vol. I, 242–93. San Francisco: W. H. Freeman.

Backus, B. T., and Sternberg, S. (1988). Attentional tradeoff across space early in visual processing: New evidence. Paper presented at the meeting of the Psychonomic Society, Chicago.

Blough, D. (1978). Reaction times of pigeons on a wavelength discrimination task. *Journal of the Experimental Analysis of Behavior, 30,* 163–167.

Broadbent, D. E. (1984). The Maltese cross: A new simplistic model for memory. *The Behavioral and Brain Sciences, 7,* 55–94.

Coltheart, M. (1985). Cognitive neuropsychology and the study of reading. In M. I. Posner and O. S. M. Marin (Eds.), *Attention and performance XI,* 3–37. Hillsdale, NJ: Lawrence Erlbaum.

Donders, F. C. ([1868] 1969). On the speed of mental processes. Translation by W. G. Koster. In W. G. Koster (Ed.), *Attention and performance II. Acta Psychologica, 30,* 412–431.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution.* New York: Basic Books.

Grayson, D. A. (1983). *The role of the response stage in stochastic models of simple reaction time.* Unpublished doctoral dissertation, University of Sydney, Australia.

Grice, G. R. (1972). Application of a variable criterion model to auditory reaction time as a function of the type of catch trial. *Perception & Psychophysics, 12,* 103–107.

Grice, G. R., Canham, L., and Boroughs, J. M. (1984). Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics, 35,* 451–463.

Johnson, E. G., and Tukey, J. W. (1987). Graphical exploratory analysis of variance illustrated on a splitting of the Johnson and Tsao data. In C. L. Mallows (Ed.), *Design, data, and analysis,* 171–244. New York: John Wiley and Sons.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* Oxford: Oxford University Press.

McCann, R. S., and Johnston, J. C. (1989). The locus of processing bottlenecks in the overlapping-tasks paradigm. Paper presented at the meeting of the Psychonomic Society, Atlanta.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

Massaro, D. W. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language, 27,* 213–234.

Meyer, D. E., Irwin, D. E., Osman, A. M., and Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed-accuracy decomposition. *Psychological Review, 95,* 183–237.

Meyer, D. E., Osman, A. M., Irwin, D. E., and Yantis, S. (1988). Modern mental chronometry. *Biological Psychology, 26,* 3–67.

Meyer, D. E., Yantis, S., Osman, A., and Smith, J. E. K. (1984). Discrete versus continuous models of response preparation: A reaction-time analysis. In S. Kornblum and J. Requin (Eds.), *Preparatory states and processes,* 69–94. Hillsdale, NJ: Lawrence Erlbaum.

Miller, J. (1988). Discrete and continuous models of human information processing: Theoretical distinctions and empirical results. *Acta Psychologica, 67,* 191–257.

Miller, J. (1990). Discreteness and continuity in models of human information processing. *Acta Psychologica, 74,* 297–318.

Miller, J., and Hackley, S. A. (N.d.). Electrophysiological evidence for temporal overlap among contingent mental processes. *Journal of Experimental Psychology: General.* In press.

Miller, J., van der Ham, F., and Sanders, A. F. (N.d.). Overlapping stage models and the additive factor method: A simulation study. Unpublished manuscript.

Miller, R. G., Jr. (1986). *Simultaneous statistical inference.* New York: Springer-Verlag.

Mosteller, F., and Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics.* Reading, Massachusetts: Addison-Wesley.

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

Osman, A., Bashore, T. R., Coles, M. G. H., Donchin, E., and Meyer, D. E. (1992). On the transmission of partial information: Inferences from movement-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 217–232.

Pashler, H., and Johnston, J. C. (1989). Chronometric evidence for central postponement in temporally overlapping tasks. *Quarterly Journal of Experimental Psychology, 41A,* 19–45.

Pieters, J. P. M. (1983). Sternberg's additive factor method and underlying psychological processes: Some theoretical considerations. *Psychological Bulletin, 93,* 411–426.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988). *Numerical recipes in C: The art of scientific computing.* Cambridge, England: Cambridge University Press.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461.

Roberts, S. (1987). Evidence for distinct serial processes in animals: The multiplicative-factors method. *Animal Learning and Behavior, 15,* 135–173.

Rumelhart, D. E., and McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

Rumelhart, D. E., and McClelland. J. L. (1986). PDP models and general issues in cognitive science. In Rumelhart, D. E. and McClelland, J. L. (Eds.), *Parallel distributed processing. Vol. 1: Foundations.* Cambridge, MA: MIT Press.

Sanders, A. F. (1980). Stage analysis of reaction processes. In G. E. Stelmach and J. Requin (Eds.), *Tutorials in motor behavior,* 331–354. Amsterdam: North-Holland.

Sanders, A. F. (1990). Issues and trends in the debate on discrete versus continuous processing of information. *Acta Psychologica, 74,* 123–167.

Schweickert, R. (1985). Separable effects of factors on speed and accuracy: Memory scanning, lexical decision, and choice tasks. *Psychological Bulletin, 97,* 530–546.

Schweickert, R. (1989). Separable effects of factors on activation functions in discrete and continuous models: *d'* and evoked potentials. *Psychological Bulletin, 106,* 318–328.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, England: Cambridge University Press.

Shaw, M. L. (1984). Division of attention among spatial locations: A fundamental difference between detection of letters and detection of luminance increments. In H. Bouma and D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes,* 109–121. Hillsdale, NJ: Lawrence Erlbaum.

Silverman, B. W. (1986). *Density estimation.* London: Chapman and Hall.

Sternberg, S. (1967). Two operations in character recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics, 2,* 45–53.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II, Acta Psychologica, 30,* 276–315.

Sternberg, S. (1971). Decomposing mental processes with reaction-time data. Invited address, Midwestern Psychological Association Annual Meeting, Detroit.

Sternberg, S. (1984). Stage models of mental processing and the additive-factor method. *The Behavioral and Brain Sciences, 7,* 82–84.

Swets, J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99,* 181–198.

Taylor, D. A. (1976). Stage analysis of reaction time. *Psychological Bulletin, 83,* 161–191.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology, 28,* 363–400.

Townsend, J. T., and Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes.* London: Cambridge University Press.

Weintz, D. A., Bickford, R. G., Bloom, F. E., and Shannahoff-Khalsa, D. S. (1983). Alternating cerebral hemispheric activity and the lateralization of autonomic nervous function. *Human Neurobiology, 2,* 39–43.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67–85.

Yantis, S., Meyer, D. E., and Smith, J. E. K. (1991). Analyses of multinomial mixture distributions: New tests for stochastic models of cognition and action. *Psychological Bulletin, 110,* 350–374.

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