

Some Constraints on Reaction-Time Distributions for Sequential Processes*

Saul Sternberg
University of Pennsylvania

ABSTRACT

It is perhaps surprising that if we know only that a process is accomplished by two or more functionally distinct subprocesses carried out in stages with stochastically independent durations, then we also know about some of the properties of the durations of that process. In particular, we know about aspects of the distributions of those durations and combinations of them. This chapter describes some of these properties. They bear on whether the reaction-time (RT) distributions from the different conditions in a factorial experiment can have the same shape, and on the conditions under which the popular ex-Gaussian distribution can describe RT distributions.

1. Sequential processes

Consider these three hypotheses:

H1 (Stages): Subprocesses **A** and **B** are stages: They operate in sequence, possibly concatenated with other operations, such that one subprocess begins when its predecessor ends.

H2 (Selective Influence): There are factors F and G , such that factor F influences only the duration T_A of stage **A**, and factor G influences only the duration T_B of stage **B**.

H3 (Stochastic Independence): Stage durations T_A and T_B are stochastically independent.

Taken together, H1, H2, and H3 constitute the S1stage Model ("SI" for stochastic independence); the more general Gstage Model embodies only H1 and H2.

A process is organized in stages if functionally distinct operations are carried out in sequential non-overlapping time periods. A stage of processing corresponds to the operation or operations that occur during one of those time periods. For a mechanism with two stages, the stream of events between stimulus and response can be cut at some time point, such that the operations before the cut are defined as stage **A**, and the operations after the cut as stage **B**, and such that the operations carried out in stages **A** and **B** are functionally distinct. Given H2, factor F acts only during the time period before the cut, influencing T_A but not T_B , while factor G acts only during the time period after the cut, influencing T_B but not T_A . Thus, if one considers the occurrence times of all operations whose durations are influenced by factor F , and similarly for factor G , these two sets of occurrence times will be contained in two different non-overlapping time periods.¹

* In T. Lachmann & T. Weis (Eds.), *Invariances in Human Information Processing* (a book honoring Hans-Georg Geissler). Taylor & Francis, In Press.

Do mental processes exist with these characteristics? Evidence consistent with the Gstage Model from behavioral measurements includes successful applications of the additive factor method.² And evidence from brain measurements has been accumulating in recent years.³ The best behavioral evidence for the S1stage Model, in my view, is provided by applications of the "summation test" (Section 2.3) by Roberts and Sternberg (1993, Section 8). Other evidence is provided when factors are found to have additive or linear effects on both $\text{var}(\text{RT})$ and $\text{mean}(\text{RT})$.⁴

2. Reaction times in an $m \times n$ factorial experiment with selective influence

2.1 Some implications of Gstage and S1stage models

Because an $m \times n$ experiment can be regarded as a concatenation of 2×2 experiments, and because two or more stages can be concatenated and treated as a single stage, we can, without loss of generality, limit ourselves to factors with two levels, and processes consisting of two stages. Let factor levels be indexed by $j = 1, 2$ for factor F , and $k = 1, 2$ for factor G . Let T_{Aj} denote the duration of stage **A** when factor F is at level F_j , etc. We then have, for the reaction time RT_{jk} at factor levels F_j and G_k , $RT_{jk} = T_{Aj} + T_{Bk}$. Notice that $RT_{11} + RT_{22} = (T_{A1} + T_{B1}) + (T_{A2} + T_{B2})$ and that $RT_{12} + RT_{21} = (T_{A1} + T_{B2}) + (T_{A2} + T_{B1})$. Because summation is associative and commutative, it follows that, for the Gstage Model, we have means additivity:

$$\overline{RT}_{11} + \overline{RT}_{22} = \overline{RT}_{12} + \overline{RT}_{21} \quad , \quad (1)$$

the basis of most applications of the additive-factor method. For the stronger S1stage Model we expect not only means additivity (Eq. 1), but also distributional additivity:

$$f_{RT_{11}+RT_{22}}(t) = f_{RT_{12}+RT_{21}}(t), \quad (2)$$

where $f(\cdot)$ is the density function, and t is reaction time. Eq. 2, in turn, implies:

$$f_{RT_{11}}(t) * f_{RT_{22}}(t) = f_{RT_{12}}(t) * f_{RT_{21}}(t), \quad (3)$$

where $*$ represents convolution.⁵

2.2 The summation property

Eqs. 2 and 3 express the *summation property*. Without tests of this property or assumptions about distributional form, the only behavioral tests of the S1stage Model of which I am aware involve evaluating the additivity of the effects of factors not only on means, but also on variances and higher cumulants, which have high sampling variability and can be influenced strongly by outliers (Ratcliff, 1979). The

-
1. Tempting as it may be to reify stages, and conflate processes with processors, Hypothesis H1 says nothing about whether the pieces of physical/anatomical machinery that carry out the operations in the two stages differ. Thus, information "transmitted from one stage to the next" doesn't necessarily move from one anatomical location to another.
 2. See, e.g., Roberts and Sternberg (1993); Sanders (1998); Sternberg (1969, 1998, 2001, 2011); and references therein.
 3. See, e.g., Anderson, Zhang, Borst, and Walsh, 2016; Cosman, Arita, Ianni, and Woodman, 2016; King and Dehaene, 2014; Schall, 2003; Schall, Purcell, Heitz, Logan, and Palmeri, 2011; Sigman and Dehaene, 2008.
 4. See Sternberg (1969, Section 5.5); Roberts and Sternberg (1993, Section 26.7); Sternberg (2016, Section 5).
 5. This was shown by Ashby and Townsend (1980, Theorem 3). For a more formal proof, see Roberts and Sternberg (1993, p. 646).

summation property implies additivity of means and higher cumulants, but also implies many other — in fact, countless other — features of the data: *Any* measures of the distribution of $RT_{11} + RT_{22}$, such as particular quantiles, or the interquartile range, or L-moments (Hosking, 1990), are predicted to be equal (within sampling error) to the same measures of the distribution of $RT_{12} + RT_{21}$.

The Gstage Model (H1 and H2) is too weak to imply the summation property. For example, the Alternate-Pathways (AP) Model, introduced by Roberts and Sternberg (1993, Section 3), is an instance of a Gstage Model that can produce additive effects on means, but violates the summation property. However, as discussed by Roberts and Sternberg (1993, Section 26.2), with an example due to Frank Norman, the SStage Model (H1, H2, and H3) is stronger than necessary for the summation property: Necessary conditions weaker than H1, H2, and H3 (but, of course, stronger than H1 and H2) have yet to be discovered. We also know (thanks to Frank Norman; see Roberts & Sternberg, 1993, Note 5) that the summation property is stronger than the combination of means additivity and variance additivity, which supports the usefulness of testing it.

2.3 Alternative tests of the summation property

How might we test the implication described by Eq. 3? Ashby and Townsend (1980) suggested directly comparing the two convolutions to test the SStage Model. To use the relation in the way they advocate, however, requires estimating density functions (Silverman, 1986) $\hat{d}_{ij}(t)$ for each of the four $\{ RT_{ij} \}$ sets, then using numerical convolution of $\hat{d}_{12}(t)$ with $\hat{d}_{21}(t)$ and $\hat{d}_{11}(t)$ with $\hat{d}_{22}(t)$, and comparing the results. To my knowledge, this method has never been applied to RT data. Density estimation of RT distributions would be especially untrustworthy (Van Zandt, 2000), given the small samples that result from the necessary partitioning of the data described in Section 2.4. The *summation test* (Roberts & Sternberg, 1993) is simpler and more direct, and has been used successfully with small data sets. The basic idea is simply to add the observed RTs for conditions 11 and 22, and for conditions 12 and 21, within levels of nuisance factors, to combine these sums across those levels, and to compare the resulting distributions. No estimation of density functions is required.

2.4 The summation test procedure

The procedure for the summation test, which is fully documented in Roberts and Sternberg (1993, Section 26.8) with examples, is as follows:

1. Partition each subject's data into what are hoped to be homogeneous subsets, i.e., within levels of the nuisance factors (experimental variations that are not of primary interest, but might influence the RT). Examples of nuisance factors may include the amount of practice (e.g., the number of previous sessions), whether or not the stimulus repeats the previous trial's stimulus, and the particular stimulus or response. Because such factors may influence the durations of more than one stage, pooling over their levels should be avoided because it may lead to spurious covariance of stage durations.
2. For each of the subsets, indexed by m , sum the elements of the cartesian products of RT_{11m} with RT_{22m} to create the $S_{11.22.m}$ set, and of RT_{12m} with RT_{21m} , to create the $S_{12.21.m}$ set. (The cartesian product of two sets of sizes n_1 and n_2 is the set of all $n_1 \times n_2$ possible pairs of their members.)
3. Before pooling these sets of sums across subsets, m , or comparing the results across subjects, adjust them by applying the same linear transformation to the members of each pair, $S_{11.22.m}$ and $S_{12.21.m}$, selecting the transformations for each pair so that the mean of their two medians and the mean of their two interquartile ranges are the same across all pairs, m . If the members of a pair of sums differ, then they will also differ after the same linear transformation is applied to them. Call these adjusted sets of sums $S_{11.22.m}^a$ and $S_{12.21.m}^a$. The principal reason for adjusting the data in this way before pooling is to maximize the chance of uncovering systematic failures of the test, which, for different pairs of distributions, seem more likely to occur at corresponding quantiles than at corresponding RTs. Also, such adjustment is likely to increase the similarity of the shapes of the pooled distributions to the shapes of the

distributions being pooled.

4. Pool each of the adjusted sets of sums over levels, m , to get $S_{11.22}^a$ and $S_{12.21}^a$ for each subject.
5. The distributions of the pooled sums $S_{11.22}^a$ and $S_{12.21}^a$ can now be compared. Because the prediction is that they should be identical within sampling error, any measures of these distributions can be used, as mentioned in Section 2.2. One possibility is to compute a pair of such distributions for each subject, determine the differences between means over subjects of the pair, for each measure of interest, and to estimate the sampling error of these differences from their between-subject variability.

2.5 Tests of the SStage Model in five data sets

To test the SStage Model, Roberts and Sternberg (1993) analyzed five data sets from four experiments (detection, identification, classification, and overlapping tasks), all of which involved two factors, each at two levels. In each case, the factors were found to have additive effects on RT means and variances. For three of the data sets (those for which individual trials data were available) the summation test was also applied, as described above; results are consistent with the SStage Model.

2.6 Sensitivity of the summation test

The AP Model (Section 2.2) is a Gstage model that is not an SStage model, with additivity of means but not variances. Based on applying the summation test to their simulations of the AP Model, Van Zandt and Ratcliff (1995) claimed that the test is insensitive. That is, they claimed that processes inconsistent with the SStage Model could satisfy the summation test.

However, their simulations were unrealistic: The interaction of the effects of factors on the variance expected from the AP Model in a 2×2 experiment is twice the product of the main effects on the RT means (Roberts & Sternberg, 1993, Eq. 15). The ratio of this product to the mean variance is one measure of the detectability of a failure of the summation test. In the data analyzed by Roberts and Sternberg (1993), this ratio ranged from 0.2 to 4.2, while it was an order of magnitude smaller (0.03) in the Van Zandt and Ratcliff simulation. Thus, relative to the main effects, the variances in their simulation were unrealistically high, rendering the summation test insensitive. Systematic testing of the sensitivity of the summation test would be desirable. When Roberts and I applied the summation test to three simulations of the AP Model that were more realistic, the deviations in each case were highly significant.⁶

3. When can the ex-Gaussian distribution describe reaction times?

3.1 The implausibility of exponentially-distributed subprocess durations

Christie and Luce (1956) and McGill (1963) suggested that the RT can be thought of as the sum of two stochastically independent random variables, corresponding to the durations of two stages, one of which is exponentially distributed. Hohle (1965) suggested that the other random variable was Gaussian and introduced the ex-Gaussian distribution, a convolution of exponential and Gaussian distributions, as a description of RTs.⁷ He argued that the duration of a decision stage might be exponentially distributed, and that because the residual latency might represent the sum of the durations of several operations with stochastically independent durations, its duration might be approximately Gaussian.⁸

To what sort of (sub)process might the exponential distribution apply? It approximates the waiting times between initiations of independent telephone calls in a large parallel network, and the times

6. These results are described, with figures and a table, by Schweickert, Fisher, and Sung (2012; sections on "Processes in series: The Summation Test", "Mixtures of processes: The Mixture Test", "Distinguishing serial processes from mixtures" and "Simulations satisfying the Mixture Test", pp. 154-163).

7. The name "ex-Gaussian" appears to have been coined by Burbeck and Luce (1982).

between decays among the innumerable atoms in a lump of radioactive material. However, because of the no-memory property unique to the exponential distribution, unless all the work is done in an instant, it cannot describe the duration of a mental process that accomplishes something: No matter how much time has elapsed, as long as processing is not yet complete, the expected remaining time needed to complete it is unchanged. (The hazard function associated with the exponential distribution is flat.) Thus, at any time point, either no work or all the work has been accomplished. The work must therefore be accomplished instantaneously. A plausible processing-time distribution must, instead, surely be one whose hazard function increases during at least some time interval.

Hohle's hope that the exponential and Gaussian components would describe the durations of functionally distinct stages of which the RT is the sum, so that different experimental factors would be associated consistently with the two stages, has not been realized (Schwarz, 2001; Matzke & Wagenmakers, 2009, especially Table 1). This inconsistency of selective influence is not surprising, given the implausibility, described above, of the duration of a mental process being exponentially distributed. Nonetheless, the ex-Gaussian distribution has recently become popular⁹ as a description of RT distributions from a wide range of experimental paradigms; and estimates of its parameters have been used to draw conclusions about underlying mechanisms, sometimes even without evaluating its goodness of fit. In some cases this distribution has been used without associating the exponential and Gaussian components with functionally distinct subprocesses (e.g., Heathcote, Popiel, & Mewhort, 1991), but in other cases, the association has been made (e.g., Penner-Wilger, Leth-Steensen, & LeFevre (2002); Shahar, Teodorescu, Usher, Pereg, & Meiran, 2014).¹⁰

The ex-Gaussian distribution was promoted by Ratcliff and Murdock (1976) (who called it the "Convolution Model", and showed that it compared favorably to two other distributions for their study-test data); by Ratcliff (1978, 1979), who used it for several memory retrieval paradigms; and by Hockley (1984), who used it for memory search and visual search data with set and display sizes from 3 to 6, as well as for other paradigms. Since then it has been used by numerous other investigators. Because of its popularity, it is of interest to ask whether the kinds of processes whose durations can be described by the ex-Gaussian distribution are constrained in some way, when they are generated by an S1stage process.

-
8. At about the same time, the ex-Gaussian distribution, also called the "lagged normal distribution", was found to describe the distribution of blood transit times through parts of the circulatory system (Basingthwaighte, Ackerman, & Wood, 1966; Davis & Kutner, 1976). Speculation about the mechanism included the idea that one part (stage) in the sequence produces Gaussian traversal delays, while another part (stage) produces exponential delays, thus adding (convolving) the two delay distributions.
 9. In PsycInfo, 136 papers published through 2015 mention "ex-Gaussian". Of these, 93% were published after 1999.
 10. Heathcote, Popiel, and Mewhort (1991) argued for using the three parameters of the ex-Gaussian distribution to describe RT distributions, noting the high sampling variability and sensitivity to outliers of the traditional moment-based measures of spread, skewness, and kurtosis. However, (1) Corresponding measures based on linear combinations of order statistics (L-moments, and the L-skewness and L-kurtosis measures derived from them (Hosking, 1990; Royston, 1992) are less sensitive to outliers than moment-based measures; (2) Such measures don't depend on the validity of a particular distributional form; (3) Parameters of the ex-Gaussian distribution do not separate distribution features of spread, skewness, and kurtosis, as changes in the single exponential parameter produce changes in all three features (as well as in the mean); (4) Sampling variability of the ex-Gaussian parameters needs to be considered; and (5) It isn't clear how to interpret the estimated parameters of a fitted distribution (such as the ex-Gaussian) if it fits poorly.

3.2 Ex-Gaussian distributions for reaction times in an $m \times n$ factorial experiment with selective influence

Again, because an $m \times n$ experiment can be regarded as a concatenation of 2×2 experiments, we can limit ourselves to factors with two levels, without loss of generality. It follows from Eq. 3 that

$$mgf_{11} \times mgf_{22} = mgf_{12} \times mgf_{21} \quad , \quad (4)$$

where mgf_{jk} denotes the moment-generating function of T_{jk} .

Now, suppose that all of the RT_{jk} are ex-Gaussian. What constraints, if any, does Eq. 4 impose on the parameters of the four ex-Gaussian distributions? There are three such constraints:

- (1) No more than one of the two factors can influence the exponential parameters $\{\tau_{jk}\}$.
- (2) The Gaussian parameters $\{\mu_{jk}\}$ must be additive. That is, the interaction $I\{\mu_{jk}\} = (\mu_{11} + \mu_{22}) - (\mu_{12} + \mu_{21}) = 0$.
- (3) The Gaussian parameters $\{\sigma_{jk}^2\}$ must be additive. That is, the interaction $I\{\sigma_{jk}^2\} = (\sigma_{11}^2 + \sigma_{22}^2) - (\sigma_{12}^2 + \sigma_{21}^2) = 0$.

A proof follows.

Let the means and variances of the Gaussian distributions be $\{\mu_{jk}\}$ and $\{\sigma_{jk}^2\}$ and the means of the exponential distributions be $\{\tau_{jk}\}$. The ex-Gaussian mgf is the product of the Gaussian mgf, $\exp(\mu t + \sigma^2 t^2/2)$, and the exponential mgf, $1/(1 - \tau t)$; namely, $\exp(\mu t + \sigma^2 t^2/2)/(1 - \tau t)$.

It therefore follows from Eq. 4 that for all t ,

$$\frac{\exp(\mu_{11}t + \sigma_{11}^2 t^2/2)}{(1 - \tau_{11}t)} \times \frac{\exp(\mu_{22}t + \sigma_{22}^2 t^2/2)}{(1 - \tau_{22}t)} = \frac{\exp(\mu_{12}t + \sigma_{12}^2 t^2/2)}{(1 - \tau_{12}t)} \times \frac{\exp(\mu_{21}t + \sigma_{21}^2 t^2/2)}{(1 - \tau_{21}t)}. \quad (5)$$

The left-hand side of Eq. 5 goes to infinity when either $t = 1/\tau_{11}$ or $t = 1/\tau_{22}$. The right-hand side goes to infinity when either $t = 1/\tau_{12}$ or $t = 1/\tau_{21}$. Because the two sides must go to infinity for the same values of t , one of the following three conditions must obtain:

- (a) All the τ_{jk} are equal, which means that neither F nor G influences τ .
- (b) $\tau_{11} = \tau_{21} \neq \tau_{22} = \tau_{12}$, which means that factor F has no influence on τ .
- (c) $\tau_{11} = \tau_{12} \neq \tau_{22} = \tau_{21}$, which means that factor G has no influence on τ .

Thus, given an SStage model with ex-Gaussian RT distributions, it is not possible for both factors (or their interaction) to influence τ .

Furthermore, in each of the three conditions above, the products of the denominators on the two sides of Eq. 3 are equal, which means that we can ignore them, and that for all t ,

$$\exp [(\mu_{11} + \mu_{22})t + (\sigma_{11}^2 + \sigma_{22}^2)t^2/2] = \exp [(\mu_{12} + \mu_{21})t + (\sigma_{12}^2 + \sigma_{21}^2)t^2/2] , \quad (6)$$

or

$$[(\mu_{11} + \mu_{22}) - (\mu_{12} + \mu_{21})]t = (1/2)[(\sigma_{11}^2 + \sigma_{22}^2) - (\sigma_{12}^2 + \sigma_{21}^2)]t^2, \quad (7)$$

or¹¹

$$I\{\mu_{jk}\} = 0, \quad (8)$$

and

11. If $at = bt^2$ for all t , we must have $a = b = 0$.

$$I\{\sigma_{jk}^2\} = 0, \quad (9)$$

where the interaction $I\{x_{jk}\} \equiv (x_{11} + x_{22}) - (x_{12} + x_{21})$. Thus, under all three conditions, (a), (b), and (c), the Gaussian means and variances must be separately additive.

These conclusions about the ex-Gaussian distribution, as well as those described below, apply whether the exponential and Gaussian components are associated with functionally distinct subprocesses, as initially proposed by Hohle (1965), or the distribution is being used purely descriptively, as proposed by Heathcote, Popiel, and Mewhort (1991).

3.3 Iterated processes and the ex-Gaussian distribution

Assume a process that generates reaction times RT_s that contain a base time (T_b) plus $s > 2$ identically distributed sequential comparison times (T_c), as in some sequential search models in which the number of elements in a display or memory set is s , and s comparisons must be made, at least on those trials on which the search target is absent:

$$RT_s = T_b + \sum_{j=1}^s T_{cj} . \quad (10)$$

Suppose that these component times are mutually independent.¹² Can the ex-Gaussian distribution be used to describe the resulting reaction-time distributions $\{RT_s\}$ for a set of three or more different s -values?

Let κ_{rb} and κ_{rc} be the r th cumulants of T_b and T_c , respectively.¹³ Eq. 10 implies that for all r ,

$$\kappa_{rs} = \kappa_{rb} + s\kappa_{rc} . \quad (11)$$

That is, each cumulant must be linear in s . The cumulant of order r of the ex-Gaussian distribution is the sum of the r th cumulants of the exponential and Gaussian distributions. For s iterations, let the exponential parameter be τ_s , and let the Gaussian parameters be μ_s and σ_s . The first four cumulants of the exponential distribution are τ_s , τ_s^2 , $2\tau_s^3$, and $6\tau_s^4$. The first four cumulants of the Gaussian distribution are μ_s , σ_s^2 , 0, and 0. It follows that the first four cumulants of the ex-Gaussian reaction-time RT_s distributions are the sums: $\kappa_{1s} = \tau_s + \mu_s$, $\kappa_{2s} = \tau_s^2 + \sigma_s^2$, $\kappa_{3s} = 2\tau_s^3$, and $\kappa_{4s} = 6\tau_s^4$, which means that $\kappa_{4s} = 3\kappa_{3s}^{4/3}$. Because κ_{4s} is a nonlinear function of κ_{3s} , they cannot both be linear in s , a contradiction. The ex-Gaussian distribution therefore cannot describe the RT distributions in this situation for more than two values of s .

It is also worth noting that because the sum of two or more exponential distributions is not exponential, if we are to identify the Gaussian and exponential components with the terms on the right-hand side of Eq. 10, then for RT_s to be ex-Gaussian, it is T_b not T_c that would have to be exponential, even for $s \leq 2$.

12. Do such processes exist? In certain memory search tasks, the linear increase in mean RT with size of the memory set is consistent with the search being sequential, and the linear increase in the RT variance is consistent with stochastic independence (Sternberg, 2016, Section 5), as is the success of other predictions (Sternberg, 2016, Section 6.3). In visual search, a linear increase in mean RT with the number of items displayed is consistent with the search being sequential (Schwarz & Miller, 2016, and references therein), but linearity of the increase of RT variance with display size for non-target trials appears never to have been carefully tested.

13. In what follows, two well-known properties of cumulants (κ_r) of order r are used: For stochastically independent random variables X and Y , $\kappa_r(X + Y) = \kappa_r[f_X(t) * f_Y(t)] = \kappa_r(X) + \kappa_r(Y)$; also, $\kappa_r(CX) = C^r \kappa_r(X)$. (See Kendall & Stuart, 1969, Volume 1, Ch. 3.)

4. Conditions for shape invariance of reaction-time distributions

4.1 The meaning and importance of shape invariance

Under what conditions do the RT distributions generated by stage models have the same shape across conditions or groups?¹⁴ Shape invariance of a set of RT distributions means that they differ by at most their means and time scales. Thus the distributions of X_1 and X_2 have the same shape if and only if there are constants C (the "scale factor") and a , such that $X_2 = a + CX_1$.

According to Ratcliff and Smith (2010, p. 90), "Invariance of distribution shape is one of the most powerful constraints on models of RT distributions. . . . That the diffusion model predicts this invariance is a strong argument in support of its use in performing process decomposition of RT data." Also, one proposal about the cognitive effects of aging is the controversial General Slowing Hypothesis: With increasing age, all the operations of the central nervous system in most or all tasks become proportionally slower (Eckert, 2011; Sleiman-Malkoun, Temprado, & Berton 2013; but see also Bashore, Wylie, Ridderinkhof, & Martinerie (2014), and Ratcliff, Spieler, & McKoon, 2000). The claim is that, in effect, with increasing age, time runs more slowly, and we should thus have shape invariance.

4.2 Two stages with selective effects on both

It follows from Eq. 3 that:

$$\kappa_{r11} + \kappa_{r22} = \kappa_{r12} + \kappa_{r21} \quad , \quad (r \geq 1), \quad (12)$$

where $\kappa_{rjk} = \kappa_r(RT_{jk})$ is the r^{th} cumulant of RT_{jk} . Let us assume that RT distributions are "well-behaved", in the sense that cumulants of (at least) orders $r = 1, 2, 3$, and 4 exist. Eq. 12 results from three assumptions: (a) *Stages*, (b) *Stochastic Independence*, and (c) *Selective Influence*. To these, let us add a fourth assumption: (d) *Shape Invariance*: The RT_{jk} distributions differ by at most means and scale factors. Whereas differences among means influence only the means of the RT_{jk} distributions and influence none of the cumulants above the first, differences among scale factors influence all of the cumulants and central moments above the first: If the distributions of two random variables, X_1 and X_2 have the same shape, with scale factor C , then

$$\kappa_r(X_2) / \kappa_r(X_1) = C^r \quad , \quad (r \geq 2). \quad (13)$$

Let the scale factor associated with RT_{jk} be $C_{jk} > 0$. It follows from Eq. 13 that

$$\kappa_{rjk} = C_{jk}^r \kappa_{r00} \quad , \quad (r \geq 2, \kappa_{r00} \neq 0). \quad (14)$$

where the $\{\kappa_{r00}\}$ are a set of constants, one for each r . With Assumption (d), Eqs. 12 and 14 then imply that:

$$C_{11}^r + C_{22}^r = C_{12}^r + C_{21}^r \quad , \quad (r \geq 2). \quad (15)$$

Given that $\kappa_{200} \neq 0$ (nonzero variances) and $\kappa_{400} \neq 0$ (nonzero kurtosis values, which, among common distributions, excludes only the Gaussian) there are only three relations among the C_{jk} that satisfy Eq. 15:

- (i) the C_{jk} are identical,
- (ii) $C_{11} = C_{21}$ and $C_{22} = C_{12}$,
- (iii) $C_{11} = C_{12}$ and $C_{22} = C_{21}$.

To prove this, use Eq. 15 with $r = 2$ and $r = 4$. For simplicity, let $C_{11} = a$, $C_{22} = b$, $C_{12} = c$, and $C_{21} = d$. Start with (A) $a^2 + b^2 = c^2 + d^2$ and (B) $a^4 + b^4 = c^4 + d^4$. Express the two sides of (A) and (B),

14. This question has been considered by Sternberg and Backus (2015), who provide an example of failure of shape invariance in data from a factorial detection experiment.

respectively, in terms of $(a + b)^2$ and $(c + d)^2$, and of $(a + b)^4$ and $(c + d)^4$. Square the two sides of the equation derived from (A), and subtract from the equation derived from (B). This gives $ab = cd$, or $a/c = d/b = k$. Substituting in (A) gives $(k^2 - 1)(c^2 - b^2) = 0$. This implies either that $k = 1$, which means that $a = c$ and $b = d$; or that $b = c$, which requires $a = d$.

Given (i), only the mean RT and none of the higher cumulants can be influenced by either factor. Given (ii), factor F can influence only the mean. Given (iii), factor G can influence only the mean. Thus, for the four RT distributions to have the same shape, at least one of the two factors can cause no more than a *shift* (a change in mean only) of the RT distribution, a highly unlikely possibility.¹⁵

It is remarkable that whereas that the four distributions of RT_{11} , RT_{12} , RT_{21} , and RT_{22} are highly likely to differ in shape, as shown above, the relations among them must be such that when they are combined in pairs, as in Eq. 3, those differences "cancel out".

4.3 Two stages with a selective effect on one

Suppose a two-stage model, with one factor F_j ($j = 1, 2, \dots$) that influences just T_A ,¹⁶ so that

$$RT_j = T_{Aj} + T_B, \quad (j \geq 1). \quad (16)$$

We then have

$$\kappa_{rj} = \alpha_{rj} + \beta_r, \quad (j \geq 1, r \geq 1), \quad (17)$$

where κ_{rj} , α_{rj} , and β_r are the r^{th} cumulants of RT_j , T_{Aj} , and T_B , respectively. Eq. 17 follows from assumptions (a), (b), and (c), in Section 4.2. Addition of the *Shape Invariance* assumption then requires

$$\kappa_{rj} = C_j^r \kappa_{r1}, \quad (j \geq 2, r \geq 2), \quad (18)$$

where $C_j \neq 1$ is the scale factor that relates RT_j to RT_1 . Combining Eqs. 17 and 18 and rearranging, we have

$$\beta_r = \frac{(\alpha_{rj} - C_j^r \alpha_{r1})}{(C_j^r - 1)}, \quad (j \geq 2, r \geq 2). \quad (19)$$

Thus, either T_B is a constant ($\beta_r = 0$, $r \geq 2$) or its distribution (which is uniquely determined up to its mean by the $\{\beta_r\}$, $r \geq 2$) is restricted by properties of the distribution of T_A , and may vary with the level of the factor F_j that is assumed to influence only T_A , a contradiction. This again implies a failure of shape invariance, given assumptions (a), (b), and (c).

4.4 Shape invariance and the ex-Gaussian distribution

Because the ex-Gaussian distribution has been fitted to numerous sets of RT data, it is interesting to ask about the conditions under which two different ex-Gaussian distributions have the same shape.

As shown in Section 3.3, the second and third cumulants of the ex-Gaussian distribution are $\kappa_2 = \tau^2 + \sigma^2$ and $\kappa_3 = 2\tau^3$. Now it is easy to show that two different ex-Gaussian distributions have the same shape if and only if $\sigma_2/\sigma_1 = \tau_2/\tau_1$: Let C be the scale factor that distinguishes the two distributions. From Eq. 14, $\kappa_{32}/\kappa_{31} = \tau_2^3/\tau_1^3 = C^3$, and $\kappa_{22}/\kappa_{21} = (\tau_2^2 + \sigma_2^2)/(\tau_1^2 + \sigma_1^2) = C^2$. The first of these implies

15. Effects of factors on mean RT are almost always associated with non-zero effects on other aspects of the distribution, including $\text{var}(RT)$. Indeed, Wagenmakers and Brown (2007) have argued for a lawful regularity in the relation between mean and variance: They claim that the standard deviations (SDs) of RT distributions increase linearly with their means.

16. This is sometimes assumed or concluded in applications of Ratcliff's (1978) Diffusion Model. See, e.g., Gomez, Perea, and Ratcliff, 2013, and Schmitz and Voss, 2012.

$\tau_2/\tau_1 = C$. Combining this with the second gives $\sigma_2/\sigma_1 = \tau_2/\tau_1$, which is thus a necessary condition for shape invariance. And because $\kappa_{r,2}/\kappa_{r,1} = \tau_2^r/\tau_1^r = C^r$ for $r > 2$, it is also a sufficient condition. See also Thomas and Ross (1980, pp. 143-144), who show that this condition is required for two ex-Gaussian distributions to be members of the same "family".

Thus a requirement for shape invariance is that any factor that influences either τ or σ should also influence the other, and in the same direction and proportion.

5. Conclusion

Behavioral and neurophysiological evidence for sequential subprocesses (stages) has been accumulating in recent years. What does this imply about the duration of a process — the reaction time — when the durations of its subprocesses are independent?

In this chapter, some of these implications are described. They include the requirement that the RT distributions from factorial experiments differ in shape, but that certain convolutions of pairs of these distributions are identical, which provides a test of the S1stage Model that requires no assumptions about distributional form. If the RT distributions in a factorial experiment are ex-Gaussian (a popular choice for fitting), then no more than one of the factors can influence the exponential parameter τ , and the Gaussian parameters μ and σ must be separately additive. We have also learned that the ex-Gaussian distribution cannot describe the RTs for more than two iterated subprocesses, as in some models of visual and memory search. The conclusions about the ex-Gaussian distribution obtain whether or not the exponential and Gaussian components are associated with functionally distinct subprocesses.

Acknowledgments

I thank David H. Krantz and M. Frank Norman for essential suggestions.

References

- Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The discovery of processing stages: Extension of Sternberg's method. *Psychological Review*, In Press.
- Ashby, F. G., & Townsend, J. T. (1980). Decomposing the reaction-time distribution: Pure insertion and selective influence revisited. *Journal of Mathematical Psychology*, *21*, 93-123.
- Bashore, T. R., Wylie, S. A., Ridderinkhof, K. R., & Martinerie, J. M. (2014). Response-specific slowing in older age revealed through differential stimulus and response effects on P300 latency and reaction time. *Aging, Neuropsychology, and Cognition*, *21*, 633-673.
- Basingthwaighe, J. B., Ackerman, F. H., & Wood, E. H. (1966). Applications of the lagged normal density curve as a model for arterial dilution curves. *Circulation Research*, *18*, 398-415.
- Burbeck, S. L. & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception & Psychophysics*, *32*, 117-133.
- Christie, L. S. & Luce, R. D. (1956). Decision structure and time relations in simple choice behavior. *Bulletin of Mathematical Biophysics*, *18*, 89-112.
- Cosman, J. D., Arita, J. T., Ianni, J. D., & Woodman, G. F. (2016). Electrophysiological measurement of information flow during visual search. *Psychophysiology*, *53*, 535-543.

- Davis, G. C., Jr. & Kutner, M. H. (1976). The lagged normal family of probability density functions applied to indicator-dilution curves. *Biometrics*, *32*, 669-675.
- Eckert, M. A. (2011). Slowing down: Age-related neurobiological predictors of processing speed. *Frontiers in Neuroscience* *5*:25, doi: 10:3389/fnins.2011.00025.
- Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked versus unmasked priming: Are they qualitatively different? *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 1731-1740.
- Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*, 340-347.
- Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 598-615.
- Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*, 382-386.
- Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society B*, *52*, 105-124.
- Kendall, M. G., & Stuart, A. (1969). *The advanced theory of statistics, Volume 1*. New York: Hafner.
- King, J.-R. & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*, 203-210.
- Matzke, D., & Wagenmakers, E.J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*, 798-817.
- McGill, W. J. (1963). Stochastic latency mechanisms. In R. D. Luce, R.R. Bush, & E. Galanter (Eds.) *Handbook of mathematical psychology, Vol. 1*. New York: Wiley. Pp. 309-360.
- Penner-Wilger, M., Leth-Steensen, C., LeFevre, J. (2002) Decomposing the problem-size effect: A comparison of response time distributions across cultures. *Memory & Cognition*, *30* 1160-1167.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59-108
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446-461.
- Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, *83*, 190-214.
- Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General*, *139*, 70 - 94.
- Ratcliff, R., Spieler, D., & McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin & Review*, *7*, 1-25.
- Roberts, S. & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.) *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive Neuroscience*. Cambridge, MA : M.I.T. Press. Pp. 611-653.
- Royston, P. (1992). Which measures of skewness and kurtosis are best? *Statistics in Medicine*, *11*, 333-343.

- Sanders, A. F. (1998). *Elements of Human Performance: Reaction Processes and Attention in Human Skill*. Mahwah, NJ: Erlbaum.
- Schall, J. D. (2003). Neural correlates of decision processes: Neural and mental chronometry. *Current Opinion in Neurobiology*, *13*, 182-186.
- Schall, J. D., Purcell, B. A., Heitz, R. P., Logan, G. D., & Palmeri, T. J. (2011). Neural mechanisms of saccade target selection: Gated accumulator model of the visual-motor cascade. *European Journal of Neuroscience*, *33*, 1991-2002.
- Schmitz, F. & Voss, A. (2012). Decomposing task-switching costs with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 222-250.
- Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, & Computers*, *33*, 457-469.
- Schwarz, W. & Miller, J. (2016). GSDT: An integrative model of visual search. *Journal of experimental psychology: Human perception and performance*, *42*, 1654-1675
- Schweickert, R., Fisher, D. L., & Sung, K. (2012). *Discovering cognitive architecture by selectively influencing mental processes*. Singapore: World Scientific.
- Shahar, N., Teodorescu, A. R., Usher, M., Pereg, M., & Meiran, N. (2014). Selective influence of working memory load on exceptionally slow reaction times. *Journal of Experimental Psychology: General*, *143*, 1837-1860.
- Sigman, M. & Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during dual-task performance. *Journal of Neuroscience*, *28*, 7585-7598.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.
- Sleimen-Malkoun, R., Temprado, J.-J., & Berton, E. (2013). Age-related dedifferentiation of cognitive and motor slowing: Insight from the comparison of Hick-Hyman and Fitts' laws. *Frontiers in Aging Neuroscience* *5*:62, doi: 10.3389/fnagi.2013.00062.
- 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. (1998). Discovering mental processing stages: the method of additive factors. In *An invitation to cognitive science: Methods, models, and conceptual issues*, D. Scarborough & S. Sternberg (Eds.). Cambridge, MA: MIT Press. Pp. 703-863.
- Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica*, *106*, 147-246.
- Sternberg, S. (2011). Modular processes in mind and brain. *Cognitive Neuropsychology*, *28*, 156-208.
- Sternberg, S. (2016). In defence of high-speed memory scanning. *Quarterly Journal of Experimental Psychology*, *69*, 2020-2075.
- Sternberg, S. & Backus, B. T. (2015). Sequential processes and the shapes of reaction time distributions. *Psychological Review*, *122*, 830-837.
- Thomas, E. A. C., & Ross, B. H. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology*, *21*, 136-152.
- Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, *7*, 424-465.

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. *Psychonomic Bulletin & Review*, 2, 20-54.

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, 114, 830-841.

CONTENTS

1. Sequential processes	1
2. Reaction times in an $m \times n$ factorial experiment with selective influence	2
2.1 Some implications of Gstage and SIstage models	2
2.2 The summation property	2
2.3 Alternative tests of the summation property	3
2.4 The summation test procedure	3
2.5 Tests of the SIstage Model in five data sets	4
2.6 Sensitivity of the summation test	4
3. When can the ex-Gaussian distribution describe reaction times?	4
3.1 The implausibility of exponentially-distributed subprocess durations	4
3.2 Ex-Gaussian distributions for reaction times in an $m \times n$ factorial experiment with selective influence	6
3.3 Iterated processes and the ex-Gaussian distribution	7
4. Conditions for shape invariance of reaction-time distributions	8
4.1 The meaning and importance of shape invariance	8
4.2 Two stages with selective effects on both	8
4.3 Two stages with a selective effect on one	9
4.4 Shape invariance and the ex-Gaussian distribution	9
5. Conclusion	10
Acknowledgments	10
References	10