

Review

A theoretical study of process dependence for critical statistics in standard serial models and standard parallel models



Ru Zhang^{a,*}, Yanjun Liu^b, James T. Townsend^b

^a Boys Town National Research Hospital, Omaha, NE, United States of America

^b Indiana University, Bloomington, IN, United States of America

ARTICLE INFO

Article history:

Received 19 March 2018

Received in revised form 5 August 2019

Available online 7 September 2019

Keywords:

Free recall

Standard serial models

Intercompletion time

Total completion time

Standard parallel models

ABSTRACT

Long before the mathematical developments that led inexorably to the development of systems factorial technology, the somewhat arduous, but arguably requisite labors which precisely defined parallel and serial architectures had begun (e.g., Townsend, 1969, 1972). Both then and now, what are now referred to as standard serial models and standard parallel models not only play an important role in psychological science, they are often what non-mathematical psychologists are (sometimes in an unschooled fashion) referring to when they bring up architectural concepts. Interestingly, two strategic and critical properties, and therefore implicit predictions of the canonical serial and parallel models have witnessed little analysis. In this article, we address three issues: (1) *Standard parallel models* predict stochastically independent processing times and therefore total completion times. There is a partially valid intuition that standard serial models will predict positive dependence among totally completion times and standard serial models based on exponential processing time do predict this (Townsend and Ashby, 1983, p. 73). This also holds if only one order of processing is possible. However, if there is a mixture of processing orders, certain distributions can predict negative dependencies on this statistic. (2) Analogously, *standard serial models* predict independent and identically distributed processing times and therefore intercompletion times. Certain examples of standard parallel models (again, based on exponential processing times) suggest that the intercompletion times tend to increase over time and stages although there has been no information on their interdependence. Interestingly, we again find that there is an inclination for standard parallel models to emulate this behavior but that this tendency can be over-ridden by dramatically increasing parallel-channel hazard functions. (3) Are there psychological questions that pertain to these distinctions and are there any experiments that have emerged to assess the latter? With regard to experimental data on these issues, it turns out that there exists some data from categorical free recall that bear on some of the behavior of intercompletion times, which we discuss. To date, we know of no investigations that explore correlations of total completion times, although such analyses appear relatively straightforward.

© 2019 Published by Elsevier Inc.

Contents

1. Introduction.....	2
1.1. Definitions of serial and parallel models.....	2
1.2. The mathematical representations of serial and parallel models.....	3
1.3. Differentiation of the serial and parallel models.....	4
1.4. The serial and parallel models' account of free recall data.....	4
1.5. The goal of this study.....	5
2. Standard two-process models.....	5
2.1. Dependence of total completion times, $n = 2$	6
2.2. Dependence of intercompletion times, $n = 2$	6
3. Standard multiple-process models.....	7
3.1. Dependence of total completion times, general n	7
3.2. Dependence of intercompletion times, general n	11

* Corresponding author.

E-mail address: ru.zhang@boystown.org (R. Zhang).

3.3. Can standard serial and standard parallel models ever mimic one another?	14
3.4. Implementation of the present theoretical results in the laboratory	15
4. Summary and conclusions	15
Acknowledgments	16
References	16

1. Introduction

An immense number of psychological tasks involve mental operations on various types of perceptual, cognitive, or action entities. Thus, questions concerning whether these operations occur in *parallel* (i.e., simultaneously), in *serial* (i.e., one at a time) or in some more complex fashion arise. Though reaching back to antiquity, the cognitive revolution beginning in the 1950s and 1960s brought renewed interest in such questions (e.g., Atkinson, Holmgren, & Juola, 1969; Egeth, 1966; Estes & Taylor, 1964; Murdock, 1971; Sperling, 1960; Sternberg, 1966).

Before beginning our technical foray, we observe that we often refer either to parallel vs. serial *systems* or parallel vs. serial *models*, which of course, are the mathematical descriptions of the material objects. Our concentration will be on response times (hereafter RTs) since that is the observable variable most frequently brought to bear on this and similar issues.

The third author's formulations, starting in the late 1960s, provided rigorous definitions of serial and parallel systems and related topics, which keyed into their most central aspects (e.g., Townsend, 1969, 1972, 1974). The serial and parallel models employed at the time in the literature tended to be not only much narrower but often defined only verbally. The rough idea of serial systems was that items would be processed in sequence and in an independent and identical-distribution, fashion. For parallel systems, all items would begin processing simultaneously and that they would be processed independently. Most often, in parallel models, it was also assumed that each item would be processed equally efficiently, no matter how many other processes were going on in parallel. This latter notion pertains to the concept of *capacity* (e.g., Townsend, 1971; Townsend & Ashby, 1978). However, we shall not be focusing on this latter assumption in this investigation, although certain statements may occasionally link up with this concept.

In our more rigorous taxonomy, these special, but important, classes of serial or parallel are now referred to as *standard serial* and *standard parallel models*. The current study focuses on these two strategic classes of models. They represent the prototypical types of serial and parallel models and were, and still are, incorrectly in our view, often depicted as the *only* members of their respective classes.

One of the theoretical challenges then and now, was that experimenters often used their data to infer one or other of the general classes of models, when in actuality, they were only testing restricted subclasses. For instance, increasing linear mean RT functions of number of items processed was taken, and still is in some quarters, as implying serial processing and excluding all parallel models. In point of fact, only special serial models make this prediction such as standard serial models. And, there exist intuitive parallel models that readily make this prediction (e.g., Townsend, 1972, 1974), in particular those suffering from *limited capacity* (e.g., Townsend, 1974; Townsend & Wenger, 2004).

There are now hundreds of experimental and theoretical works on the serial and parallel question (for recent reviews, e.g., Algom, Eidels, Hawkins, Jefferson, & Townsend, 2015; Townsend & Wenger, 2004; Townsend, Wenger, & Houpt, 2018; Townsend,

Yang, & Burns, 2011). The rigorous theoretical scaffolding constructed by Townsend and colleagues led not only to understanding of where such issues of parallel vs. serial processing cannot be experimentally adjudicated, but also, more happily, to quite powerful theory-driven methodologies for testing the two types of process arrangements (e.g., Townsend, 1972, 1974; Townsend & Ashby, 1983; Townsend & Nozawa, 1995). Among the most powerful approaches is the topic of this volume, Systems Factorial Technology (SFT; see especially Little, Altieri, Fific, & Yang, 2017). Although the present article does not specifically appeal to SFT, it certainly lies squarely in its tradition, which began back in the 1960s with the aforementioned rigorous taxonomy of process models and scrupulous exploration of parallel and serial systems' properties. It is true that dependencies among subsystems (defined as "processors" below) can be deleterious to the workings of SFT. That is, inter-channel processor dependencies can interfere with the critical assumption of "selective influence" (Townsend, 1984). This type of perturbation is known as "indirect non-selective influence". Progress on such contamination has been made by Townsend and Thomas (1994), Townsend and Ashby (1983, Chapter 12) and Dzharfarov (1999). There are still important unanswered questions in that direction, but these lie outside the scope of our present enterprise. Theoretical methodologies closely related to SFT include many works by Schweickert, Dzharfarov and colleagues (see e.g., Schweickert, Fisher, & Sung, 2012, for a broad-spectrum review and discussion). Other important approaches to questions of mental architecture are Anderson (2013) and Newell (1994).

1.1. Definitions of serial and parallel models

We need to recall the rigorous definitions of serial vs. parallel models. Our strategy obeys the precept employed in our first rigorous analyses of information processing systems, that of representing architecture and stopping rules by way of the fundamental building-bricks called "intercompletion times". These will be precisely defined in what follows. The arrangement and relative speeds describing intercompletion times can handle almost any issues concerning the characteristics of information processing including the more complex architectures known as PERT networks (e.g., Schweickert, 1978; Schweickert et al., 2012).

The serial models assume all the processes are executed in an end-to-end manner (Fig. 1(a)). If n processes are in the model, there are $n!$ ways to arrange the order of them. The parallel models assume that every process starts simultaneously (Fig. 1(b)) but they can terminate at different moments. There are also, like the serial case, $n!$ ways to terminate the n processes in the parallel models. For informal discussion the term *channel*, instead of *process*, will sometimes be associated with parallel systems and models. However, the reader should refer to the more precise mathematical definitions and developments if our meaning is ever unclear.

In this article, we distinguish between the terms *processor j* and the *j*th order of processor. The "j" in the term "processor j" represents the name or identity of a subsystem process in operation whereas "jth order of processor" refers to the position in the processing order of a processor. Furthermore, we refer to the *processing time* as the actual time to execute a processor, which will, of course be distinct for serial and parallel processors.

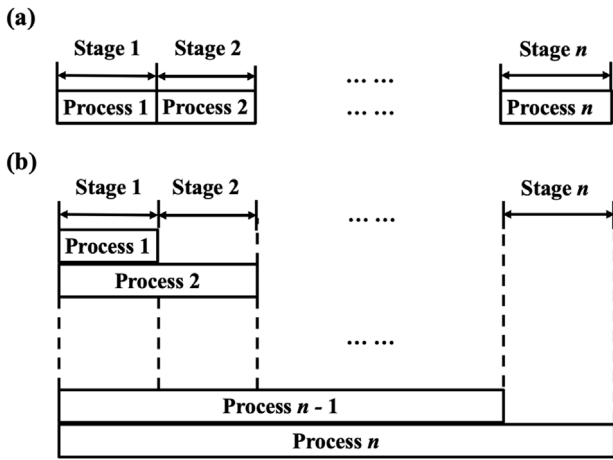


Fig. 1. Graphical representations of (a) a serial model and (b) a parallel model.

For brevity, the term “process” is employed in our figures but the reader should always think of each process in a figure as identifying a specific processor.

Thus, the variable z_j denotes the true processing time for item associated with processor j , where $j = 1, \dots, n$. The “ j ” in the term “the j th order of processor” indicates that there are $j - 1$ processors that have finished before that processor. In Fig. 1, for simplicity processor j and the j th order of processor are the same. But please keep in mind, processor j might be the i th order of processor, where $i \neq j$. Thus, in Fig. 1 processor 1 might finish as the i th ($i = 1, 2, \dots, n - 1, n$), whether in a serial or a parallel system.

Stage. Stage j spans the interval from the moment the $(j - 1)$ th processor to be completed to the moment the j th processor to finish is complete. Sometimes, the term “stage” is used only in serial models, but we also employ it to act as this kind of descriptive statistic.

Nonetheless, for serial models, the duration of stage j is identical to the duration of the j th intercompletion time. In Fig. 1(a), processor 1 occupies stage 1, processor 2 occupies stage 2, ..., and processor n is functioning (only) during stage n .

In contrast, for parallel models, a processor will, except for stage 1, include several stages. For instance, processor 2 in Fig. 1(b) includes stage 1 and stage 2 and processor n includes stage 1, stage 2, to stage n .

Intercompletion time. As intimated above, the *intercompletion time* T_j (also named the j th intercompletion time or j th ICT) is the time random variable that is spent for stage j whether serial or parallel. Again, for the serial models, each processing time is an ICT: In Fig. 1(a),

$$T_1 = z_1, T_2 = z_2 \dots T_n = z_n.$$

In contrast, the ICT is only part of a processing time for the parallel models for all but the very first stage: In Fig. 1(b),

$$T_1 = z_1, T_2 = z_2 - z_1 \dots T_n = z_n - z_{n-1}.$$

Total completion time for a processor. The *total completion time* \mathbb{T}_j (TCT) is the time that is consumed from the onset of processing by the system, to the moment that processor j is complete. For the serial model in Fig. 1(a),

$$\mathbb{T}_1 = z_1, \mathbb{T}_2 = z_1 + z_2, \dots, \mathbb{T}_n = z_1 + z_2 + \dots + z_n.$$

But, for the parallel model in Fig. 1(b),

$$\mathbb{T}_1 = z_1, \mathbb{T}_2 = z_2, \dots, \mathbb{T}_n = z_n.$$

To sum up so far, z_i will be a TCT in a parallel system but an ICT in a serial system. \mathbb{T}_j can also be written as the sum of ICTs up to processor j .

Total completion time for a stage. The TCT \mathbb{S}_j is the time that is consumed from the onset of the system, to the moment that stage j is complete. \mathbb{S}_j can also be written as the sum of ICTs up to stage j .

We summarize relationships between TCT for a stage \mathbb{S}_j , TCT for a processor \mathbb{T}_j , ICT T_j and processing time z_j as following. For the serial model in Fig. 1(a),

$$\mathbb{S}_1 = \mathbb{T}_1 = T_1 = z_1,$$

$$\mathbb{S}_2 = \mathbb{T}_2 = T_1 + T_2 = z_1 + z_2,$$

...

$$\mathbb{S}_n = \mathbb{T}_n = T_1 + T_2 + \dots + T_n = z_1 + z_2 + \dots + z_n.$$

For the parallel model in Fig. 1(b),

$$\mathbb{S}_1 = \mathbb{T}_1 = T_1 = z_1,$$

$$\mathbb{S}_2 = \mathbb{T}_2 = T_1 + T_2 = z_2,$$

...

$$\mathbb{S}_n = \mathbb{T}_n = T_1 + T_2 + \dots + T_n = z_n.$$

Again, to be sure we're all on the same page, note that the intercompletion time T_j and the total completion time for a stage \mathbb{S}_j are defined with respect to the stage in the processing order, whereas the total completion time for a processor \mathbb{T}_j is defined with respect to the identity of a processor. And, again in this article, we frequently refer the term *processing time* (the same as the “actual processing time” in Townsend & Ashby, 1983) the actual time to execute a processor. To stamp in this concept, for a serial model, processing time is equivalent to intercompletion time. For a parallel model, processing time is equivalent to total completion time.

1.2. The mathematical representations of serial and parallel models

Both serial models and parallel models can be represented mathematically. With the foregoing definitions in hand, we shall forthwith employ ICT to designate intercompletion time and TCT to refer to total completion times.

By a natural convention, usually a serial model is associated with ICTs and a parallel model with TCTs, although a serial model could be defined in terms of TCTs and a parallel model could be defined in terms of ICTs (for theoretical purposes, it is often desirable to express both classes in terms of ICTs as in Townsend & Ashby, 1983).

A serial model can be written as the product of the probability of a certain serial order of processing and the joint density function of ICTs conditioned on that order.

$$P(I) f_s(T_1 = t_1, \dots, T_n = t_n | I = (i_1, \dots, i_n)).$$

t_1, \dots, t_n are realizations of T_1, \dots, T_n , $(i_1, \dots, i_n) \in Perm(n)$, where $Perm(n)$ is the set of all permutations of the naturals from 1 to n , and $P(I)$ is the probability of a particular permutation $I = (i_1, \dots, i_n)$. For serial models, permutation I means that the model starts with processor i_1 and is connected to the onset of processor i_2 after processor i_1 is complete, and so on. In contrast, parallel models can be most naturally written as the joint density function of TCTs of processors:

$$f_p(\mathbb{T}_1 = \tau_1, \dots, \mathbb{T}_n = \tau_n; I),$$

where τ_1, \dots, τ_n are realizations of $\mathbb{T}_1, \dots, \mathbb{T}_n$. For parallel models, permutation I means that all the processors start simultaneously but processor i_1 terminates first, processor i_2 terminates second, and so on.

1.3. Differentiation of the serial and parallel models

Without imposing additional assumptions, serial models and parallel models can perfectly mimic each other in many experimental situations. Then, the experimenter finds it impossible to tell them apart (e.g., Townsend & Ashby, 1983, Chapter 14; Hout, Townsend, & Jefferson, 2017). In order to overcome this problem, several candidate assumptions were raised by scientists.

Selective influence (Sternberg, 1969) is the most widely used one. It states that manipulation of each factor only influences the processor that is associated with that factor. In serial models, the outcome is that mean RTs can be shown to be additive functions of the set of factors. Observe that this statement holds even if the serial model falls outside the class of standard serial models. Over the years, the assumption of selective influence was generalized and refined by Townsend and others (e.g., Dzhaferov, 2003; Schweickert & Townsend, 1989; Townsend, 1984; Townsend & Schweickert, 1989; Townsend & Thomas, 1994).

In fact, SFT (Townsend & Nozawa, 1995) was developed to differentiate parallel models from serial models based on that assumption. One can diagnose the nature of the process arrangements according to the sign of the mean interaction contrast of RTs (e.g., Schweickert, 1978; Sternberg, 1969; Townsend, 1984; Townsend & Schweickert, 1985).

Moreover, much more precise assays were made possible by the extension of contrast functions to survival functions of RTs (Townsend & Nozawa, 1995). SFT has been applied and extended and further explored by a plethora of researchers (e.g., Dzhaferov, Schweickert, & Sung, 2004; Little et al., 2017; Schweickert, Giorgini, & Dzhaferov, 2000; Yang, Fific, & Townsend, 2013; Zhang & Dzhaferov, 2015).

One limitation of SFT is that it requires application in a *complete factorial design*, in which each factor has at least two levels (low salience vs. high salience that result in long processing time vs. short processing time). Not every feature or dimension of interest can fulfill this requirement. For instance, the processing time of red color may be neither faster nor slower than green.

Another candidate assumption is that of *within stage independence*, which states that unfinished parallel processors are independently executed within each stage. Since there is only one processor in each stage in a serial model, this assumption is only applicable in the class of parallel models. Although within stage independence is an important characteristic to know about, it turns out that within stage dependent models can be mathematically transformed to within stage independent models (e.g., see Rao, 1992, pp. 162–163). Thus, within stage independent parallel models and within stage dependent parallel models cannot be discriminated in the absence of direct observability of the within stage dependencies. The next paragraph is critical to understanding the remainder of our developments and is thereby italicized though it includes concepts already introduced.

In this article, the axioms concerning processing time independence are absolutely central. Recall that processing time is defined as the actual time to execute a processor. For the serial models, this assumption is equivalent ICT independence since the actual processing times are equivalent to the ICTs. In contrast, for the parallel models, processing time of an item or channel is equivalent to the time spent from the very initiation of the model until an individual channel is finished. As introduced earlier, this is a statistic known generically as the TCT for a processor. Independence of processing times in parallel models is tantamount to independence of the TCTs for individual processors. In addition, we assume each processing time is identically distributed. By assuming independently and identically distributed (i.i.d) processing times, serial models and parallel models are termed standard serial models and standard parallel models, respectively (e.g., Algom et al., 2015; Townsend et al., 2018).

In order to investigate the complete dependency structures, we must consider the joint distribution of all component ICTs (TCTs, etc.). This is tantamount to assuming exhaustive processing but that is a simple accident of our need for the overall joint distributions rather than a wish to focus on that particular stopping rule.

1.4. The serial and parallel models' account of free recall data

Many experimental paradigms are traditionally used to systematically study the serial–parallel issue. Free recall is one of them, though far from the most popular. In this paradigm, subjects are instructed to recall words that belong to a semantic category from their long-term memory (Bousfield & Sedgewick, 1944; Bousfield, Sedgewick, & Cohen, 1954), for instance, naming as many cities in the United States as they could remember. The words are reported successively. It is found that the time interval between two successive responses, that is the ICT, increases as more responses are generated (Lohnas, Polyn, & Kahana, 2011; Murdock & Okada, 1970; Patterson, Meltzer, & Mandler, 1971; Pollio, Kasschau, & DeNise, 1968; Pollio, Richards, & Lucas, 1969; Polyn, Norman, & Kahana, 2009; Sederberg et al., 2006; Sederberg, Miller, Howard, & Kahana, 2010).

Various serial and parallel models have been proposed to interpret the behavior of the data that is observed in this paradigm. Here we introduce several important ones. McGill contributed an influential chapter on stochastic processes in psychology, to the 1963 Volume 1 of the Handbook of Mathematical Psychology (McGill, 1963). He accounted for the general temporal characteristics of Bousfield and Sedgewick's (1944) data based on a serial model. His model assumes that only one item could be sampled from a search set and inspected at any time. All the relevant items are assumed to be chosen with the equal chance at each draw. After each draw, the subject examines if the item is a member of the specified category and if the item is not recalled yet. The amount of time for each draw and subsequent check is assumed to be exponentially distributed with the same rate parameter, namely equal accessibility. Intriguingly, McGill's (1963) serial model is mathematically identical to the standard parallel model with exponential processing times.

Vorberg and Ulrich (1987) subsequently generalized McGill's model according to the assumption of unequal accessibility. By allowing some items more easily to be accessed from memory than the others (Shiffrin, 1970), the generalized model removes some minor discrepancies between the data and McGill's model in predicting the number of generated items by a certain time moment. The stochastic representation of Vorberg and Ulrich (1987)'s serial model is then found to be

$$P(I)f_s(T_1 = t_1, \dots, T_n = t_n | I = (i_1, \dots, i_n)) \\ = \left(\prod_{j=1}^n \frac{u_{i_j}}{\sum_{l=j}^n u_{i_l}} \right) \left[\prod_{j=1}^n \left(\sum_{l=j}^n u_{i_l} \right) \exp \left(- \sum_{l=j}^n u_{i_l} t_j \right) \right],$$

where u_{i_j} stands for the rate parameter for processor (or item) i_j and n is the number of recallable target items within the search set. The model predicts that the rate parameter of the j th ICT equals the sum of rate parameters of processors that have not been executed, that are the $j + 1$ th order of processor to the n th order of processor. Note that when the equal accessibility assumption is imposed ($u = u_{i_1} = \dots = u_{i_n}$), the ICT distribution does not depend on the recall order any more and the conditional joint density function is reduced to

$$f_s(T_1 = t_1, \dots, T_n = t_n | I = (i_1, \dots, i_n)) \\ = \prod_{j=1}^n (n - j + 1) u \exp \left[- (n - j + 1) u t_j \right],$$

which is the original McGill (1963) model. Note by observing the above equation, McGill (1963)'s model indicates that the ICTs, or equivalently processing times, are not i.i.d. Vorberg and Ulrich (1987) also derived the stochastic representation of the counterpart parallel model with the assumption of unequal accessibility:

$$f_p(\mathbb{T}_1 = \tau_1, \dots, \mathbb{T}_n = \tau_n; I = (i_1, \dots, i_n)) = \prod_{j=1}^n u_{i_j} \exp(-u_{i_j} \tau_{i_j}).$$

Observe that the TCTs are stochastically independent of each other in the counterpart parallel model. By imposing the equal accessibility assumption ($u = u_{i_1} = \dots = u_{i_n}$), the above equation is reduced to

$$f_p(\mathbb{T}_1 = \tau_1, \dots, \mathbb{T}_n = \tau_n; I = (i_1, \dots, i_n)) = \prod_{j=1}^n u \exp(-u \tau_{i_j})$$

which is McGill (1963)'s model in the parallel model's representation. Observe it is also the mathematical representation of the standard parallel model with exponential processing times.

Following McGill's model, Rohrer and Wixted (1994) derived a function for the mean ICT changes as more items are reported by the subjects:

$$\bar{T}_j = \frac{1}{u(n-j)},$$

where \bar{T}_j is the mean of the j th ICT. This equation reflects that an ICT is inversely proportional to $n - j$. Thus, the last ICT of a four-item recall should equal the last ICT of a nine-item recall.

Rohrer and Wixted (1994) conducted experiments by asking the subjects to recall the words studied earlier. By manipulating the size of the study word list or/and the presentation time for each word, they found that the hyperbolic ICT growth reflected by the above equation fitted the data well. As mentioned earlier, McGill's serial model is identical to the standard parallel model with exponential processing times. Thus, Rohrer and Wixted (1994)'s study supports the independent and identical processing time assumption if one considers the recall a parallel process.

1.5. The goal of this study

Our work complements the work on differentiating the serial models and parallel models. We characterize the foundational distinctions between standard serial models (a special class of serial models) and standard parallel models (a special class of parallel models). This goal leads us directly to investigate the dependence of TCTs, the actual processing times of parallel systems as well as the dependence of ICTs, the actual processing times of serial systems. As we have mentioned, a primary axiom of a standard serial model is that the ICTs (i.e., the processing time) are independent. In contrast, for a standard parallel model, the TCTs (i.e., the processing times) are independent. Thus, our mission in this work, is to explore the ICT dependence characteristics of standard parallel models and the TCT dependence characteristics of standard serial models. To be more specific, we investigate the behavior of (conditional) distributions, of the TCT of processors and the ICT in complete generality, that is, without assuming any particular form for the distributions of processing times.

The pioneers (McGill, 1963; Rohrer & Wixted, 1994; Vorberg & Ulrich, 1987) developed serial models or/and parallel models to account for temporal characteristics observed in free recall experiments. Some model they developed can be considered as a restricted version of the standard parallel model. We explore

further by examining under what condition the theoretically derived behavior from the (unrestricted) standard parallel model is consistent with the empirical findings that the ICT grows as a function of output position.

In a free recall task, one has to exhaustively report all the items that belong to a specific category from memory. This type of stopping rule is traditionally termed as "exhaustive" (Townsend & Nozawa, 1995). The standard serial models and standard parallel models are associated with this rule: All the processors in the models have to be completed to conclude a model. If a model is terminated once one processor out of the n processors is fully executed, this type of stopping rule is traditionally termed as "self-terminating". In this article, all the results are irrelevant to "self-terminating".

In a previous publication (Zhang, Liu, & Townsend, 2018), our treatment was limited to two channels or stages in operation. In the current study, we extend our conclusions to the case of arbitrary n . In order to facilitate comprehension of the general case, we provide a brief review of findings for the two-process models in the next section. All the theorems, corollaries, and lemmas in this section were reported in our earlier publication (Zhang et al., 2018). The readers can access the mathematical proofs and computational simulations associated with those theorems, corollaries, and lemmas from that chapter.

2. Standard two-process models

Suppose there are only two processors a and b in the models to realize standard two-process serial models and the comparable standard two-process parallel models.

Let us denote the processing times of a and b as z_a and z_b (recall that they are i.i.d, whether in a serial or a parallel model) and the density function for each as f . The corresponding distribution function is labeled as F . The survival function, the hazard function, and the cumulative hazard function are represented respectively as

$$S(x) = 1 - F(x),$$

$$h(x) = \frac{f(x)}{S(x)},$$

$$H(x) = \int_0^x h(x) dx = -\ln[S(x)],$$

where x denotes a temporal variable. The readers should pay attention to the notation. In this paper S always denotes the survival function.

Standard two-process serial models. Since two processors are under consideration, the model can be decomposed into two stages. If processor a is executed earlier than processor b , then processor a is stage 1. If processor a is executed later than processor b , then processor b is stage 1 (Fig. 2).

As defined earlier, the ICT T_1 is the time that is spent processing in stage 1 and T_2 is the time that is spent processing in stage 2. So, for Case I, $T_1 = z_a, T_2 = z_b$ and for Case II, $T_1 = z_b, T_2 = z_a$. It is apparent that T_1 and T_2 are i.i.d as z_a and z_b are assumed i.i.d. The TCT \mathbb{T}_a is the time that is consumed from the onset of the model to the moment that processor a is completed. The TCT \mathbb{T}_b is the time that is consumed from the onset of the model to the moment that processor b is completed. Therefore, for Case I, $\mathbb{T}_a = T_1 = z_a, \mathbb{T}_b = T_1 + T_2 = z_a + z_b$ and for Case II, $\mathbb{T}_a = T_1 + T_2 = z_a + z_b, \mathbb{T}_b = T_1 = z_b$.

Standard two-process parallel models. Since two processors are under consideration (see an example in Fig. 3), the TCT for processor a and the TCT for processor b are

$$\mathbb{T}_a = z_a,$$

$$\mathbb{T}_b = z_b,$$

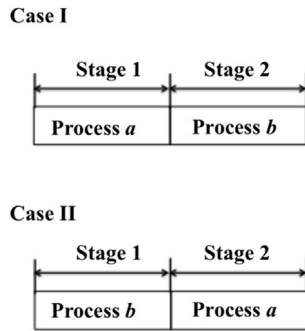


Fig. 2. Possible processor arrangements of a standard two-process serial model. From “A Theoretical Study of Process Dependence for the Standard Two-Process Serial Models and Standard Two-Process Parallel Models”, by R. Zhang, Y. Liu, and J.T. Townsend, in T. Lachmann and T. Weis (Eds.), *Invariances in Human Information Processing* (pp. 117–142), 2018, New York, NY: Taylor & Francis Group. Source: Reprinted with permission. © 2018 Taylor & Francis Group.

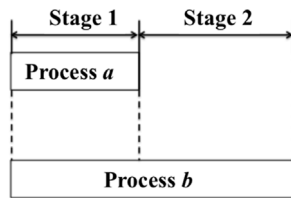


Fig. 3. A standard two-process parallel model. From “A Theoretical Study of Process Dependence for the Standard Two-Process Serial Models and Standard Two-Process Parallel Models”, by R. Zhang, Y. Liu, and J.T. Townsend, in T. Lachmann and T. Weis (Eds.), *Invariances in Human Information Processing* (pp. 117–142), 2018, New York, NY: Taylor & Francis Group. Source: Reprinted with permission. © 2018 Taylor & Francis Group.

respectively. Please note that Fig. 3 is an exemplar representation of a standard two-process parallel model in which processor a is faster than processor b . With some non-zero probability, processor a will be slower than processor b as z_a and z_b are i.i.d. The ICTs in Fig. 3 can be represented as

$$T_1 = \mathbb{T}_a = z_a,$$

$$T_2 = \mathbb{T}_b - \mathbb{T}_a = z_b - z_a.$$

2.1. Dependence of total completion times, $n = 2$

As observed earlier, the TCT of a parallel channel (item, etc.) is its processing time and perforce, in a standard parallel model is independent of the processing times of the other channels. An intuition about the potential comparable dependencies in a serial model is the following: condition TCT for processor “ b ” on that for “ a ”. Now, if “ b ” is done first then, since the TCT for “ a ” is the sum of the two processing times for “ a ” plus that of “ b ”, the probability that the TCT for “ b ” is less than τ , given that the sum is already $< \tau$, must be 1. If “ a ” is done first by time τ , the probability that “ b ” also gets done by τ is greater than its marginal probability. Thus, this qualitative intuition suggests that the TCT in a serial model should be positively dependent. We will learn that this reasoning is faulty in general but correct when only one processing order is allowed.

We move on to perform the actual required computations, comparing the distribution function of \mathbb{T}_b conditional on \mathbb{T}_a versus the marginal distribution function of \mathbb{T}_b . That is

$$P(\mathbb{T}_b \leq \tau | \mathbb{T}_a \leq \tau) - P(\mathbb{T}_b \leq \tau). \quad (1)$$

If it is always positive then we conclude that the TCTs in this case are positively dependent in a strong distributional sense, and conversely if the difference is negative. Please note that (1) considers two possible permutations as illustrated in Fig. 2.

It was proven by Townsend and Ashby (1983, Page 73–74), if the processing times z_a and z_b (or the ICTs T_1 and T_2) in a two-process serial model are i.i.d and follow exponential distributions, then (1) > 0 for $\tau > 0$. This result is compatible with the above intuition. However, we recently found (1) > 0 does not hold for all processing time distributions (Zhang et al., 2018).

Theorem 1. For a standard two-process serial model, $P(\mathbb{T}_b \leq \tau | \mathbb{T}_a \leq \tau) - P(\mathbb{T}_b \leq \tau)$ can be either positive or nonpositive for $\tau > 0$.

Corollary 2 states if only one permutation is allowed in the investigated system, dependence of TCTs is non-negative, that is (1) ≥ 0 . We also compute the covariance of TCTs. The result is presented in Lemma 3.

Corollary 2 and Lemma 3 correspond to our original intuitions and the Townsend and Ashby (1983) theorem when the distributions are gamma (Erlang). The covariance result in Lemma 3 is especially pleasing.

Corollary 2. For a standard two-process serial model, $P(\mathbb{T}_b \leq \tau | \mathbb{T}_a \leq \tau) - P(\mathbb{T}_b \leq \tau)$ is nonnegative for $\tau > 0$ if only the Case I arrangement or only the Case II arrangement is allowed in the investigated system.

Lemma 3. For a standard two-process serial model, if only the Case I arrangement or only the Case II arrangement is allowed, $\text{Cov}(\mathbb{T}_a, \mathbb{T}_b) = \text{Var}(T_1) > 0$.

We know by the very definition of standard parallel models, that the TCTs are independent. However, we include the obvious statements for ease of reference. It is therefore listed that (1) = 0 and this statement is presented in as Theorem 4. We also compute the covariance of TCTs for standard two-process parallel models. The result is presented in Lemma 5.

Theorem 4. For a standard two-process parallel model, $P(\mathbb{T}_b \leq \tau | \mathbb{T}_a \leq \tau) - P(\mathbb{T}_b \leq \tau) = 0$ for $\tau > 0$.

Lemma 5. For a standard two-process parallel model, $\text{Cov}(\mathbb{T}_a, \mathbb{T}_b) = 0$.

Standard two-process serial models and standard two-process parallel models can be differentiated according to Theorems 1 and 4: $P(\mathbb{T}_b \leq \tau | \mathbb{T}_a \leq \tau) - P(\mathbb{T}_b \leq \tau)$ cannot always be zero for a standard two-process serial model; while as for a standard two-process parallel model, the function maintains zero along the axis of τ . One can also differentiate the two models according to Lemmas 3 and 5.

2.2. Dependence of intercompletion times, $n = 2$

The ICTs, and therefore the processing times, in a standard serial model are assumed i.i.d. Therefore, the empirical finding in free recall tasks that as the number of stages already completed increases, the ICTs increase, cannot be accounted for by standard serial models. In contrast, standard parallel models can account for this phenomenon as noted by McGill (1963) and Vorberg and Ulrich (1987). The intuition of course, is that the later stages included fewer and fewer parallel processes still to complete and therefore the probability that the minimum time for one of these remaining processors to finish inevitably lengthens. However, it is so far unknown as to whether this behavior is characteristic of

all standard parallel models. We shall learn that it is not. First, however, we will investigate a related, but distinct question.

Without loss of generality, we label the processor completed earlier processor a and the other is labeled as processor b in a standard two-process parallel model. Recall that the processing times z_a and z_b , or equivalently \mathbb{T}_a and \mathbb{T}_b , are assumed i.i.d. Now let us label the ICTs for stage 1 and stage 2 as T_a and T_b , where

$$T_a = \mathbb{T}_a = z_a,$$

$$T_b = \mathbb{T}_b - \mathbb{T}_a = z_b - z_a.$$

One topic of interest is how the ICTs act in later stages as a function of the magnitude of the earlier stages. For $n = 2$, this simply suggests we investigate the likelihood that the second stage ICT T_b , is not completed by time t , given that the first stage is entirely completed. We thus explore the survival function of the ICT T_b conditional on the processing time of stage 1: $P(T_b > t | z_b > z_a)$, where $t > 0$. Interestingly, it is found that the behavior of $P(T_b > t | z_b > z_a)$ depends on the hazard function of processing time h .

Lemma 6. For a standard two-process parallel model, if the hazard function h for an arbitrary channel is non-increasing, then $P(T_b > t | z_b > z_a)$ is non-decreasing as T_a is increased.

This result begins to tie the behavior of subsequent ICT times with the hazard function. The finding that they stochastically lengthen as a function of the previous finishing time, if the hazard function is non-increasing, makes intuitive sense.

A separate issue, and one more directly related to the question of whether successive ICTs tend to grow longer, is “How does the later stage duration (stage 2) compare with the earlier stage duration (stage 1) in a standard two-process parallel model?” Let us denote the ratio of the hazard functions:

$$\alpha(t, T_a + t) = \frac{h(T_a + t)}{h(t)}.$$

The survival function at stage 1 in the standard two-process parallel model is the product of two survival functions $S^2(t)$. The survival function at stage 2 is

$$P(T_b > t | z_b > z_a)$$

$$= P(T_b + T_a > t + T_a | z_b > z_a)$$

$$= P(z_b > t + T_a | z_b > z_a)$$

$$= \frac{S(T_a + t)}{S(T_a)}.$$

We then investigated $S^2(t)$ vs. $\frac{S(T_a+t)}{S(T_a)}$. If the survival function from stage 1 to stage 2 is increasing, that is $S^2(t) - \frac{S(T_a+t)}{S(T_a)} < 0$, this trend is then consistent with the empirical finding that the ICT grows as the number of stages grows, in a strong distributional sense. **Theorem 7** provides an exact condition for the survival function to increase from stage 1 to 2. **Corollary 8** states that standard two-process parallel models with a concave or linear cumulative hazard function $H(t)$ predict an increasing survival function from stage 1 to stage 2, for arbitrary t .

Theorem 7. In a standard two-process parallel model, if $\alpha(t, T_a + t) < 2$, then $S^2(t) - \frac{S(T_a+t)}{S(T_a)} < 0$ so that the survival function of ICT is increasing from the first stage to the second stage; if $\alpha(t, T_a + t) \geq 2$, then $S^2(t) - \frac{S(T_a+t)}{S(T_a)} \geq 0$ so that the survival function of ICT is non-increasing from the first stage to the second stage.

Thus, we have the intriguing and reasonable prediction that any standard two-process parallel model whose hazard function ratio α is less than the number 2 (i.e., $\alpha < 2$) will evidence increasing conditional survivor functions and hence result in the

longer ICT in the second stage than the first stage. This result has an elegant consequence that can be expressed in terms of the integrated hazard functions, as visited in **Corollary 8**.

Corollary 8. For a standard two-process parallel model, (i) if the cumulative hazard function $H(t)$ is concave or linear, then $S^2(t) - \frac{S(T_a+t)}{S(T_a)} < 0$; (ii) if $H(t)$ is convex, then the sign of $S^2(t) - \frac{S(T_a+t)}{S(T_a)}$ is uncertain.

The quick interpretation is that anything as slow or slower than an exponential (recall that an exponential distribution has a linear $H(t)$), and therefore constant, hazard function will imply stochastically increasing ICTs. If $H(t)$ is convex, indicating an increasing $h(t)$, then more information must be garnered. The upshot is, as intuition may suggest, that standard parallel models appear to tend to produce increasing ICTs, but that inclination can be defeated by dramatically increasing hazard functions.

3. Standard multiple-process models

We now proceed to generalize the theorems for standard two-process models to standard multiple-process models. All the theorems, corollaries, and lemmas in this section are new. Suppose there are processors $1, \dots, n$ in the model. Recall that $1, \dots, n$ in the term “processors $1, \dots, n$ ” represent the identities of the processors not the positions in the processing order. The corresponding processing times z_1, \dots, z_n are assumed to be i.i.d with the density function f . The corresponding distribution function, survival function, hazard function, and the cumulative hazard function are labeled as F, S, h , and H , respectively. At this point, we do not have a completely analytic proof for arbitrary n . However, we will write down the general expressions and then pick an arbitrary value of n with which to perform the pertinent computations.

3.1. Dependence of total completion times, general n

There are $n!$ ways to arrange the n processors in the standard n -process serial models. If we were to assay the behavior of models where the identity or location of a processor was associated with distinct probability distribution, that would necessitate a more complex notation. However, due to the assumption of i.i.d random variables, we can invoke a much-simplified set of symbols (see **Townsend & Ashby, 1983**, Chapter 15; **Townsend et al., 2018**; or **Haupt et al., 2017**, for the more general situation).

Thus, **Fig. 4** presents two particular cases where processors $1, \dots, j$ are processed earlier than processors $j + 1, \dots, n$ (Case I) and processors $1, \dots, j$ are processed later than processors $j + 1, \dots, n$ (Case II). According to the earlier definition, \mathbb{T}_j is the time that is consumed from the onset of processing to the moment that processor j is complete, where $1 \leq j \leq n$. S_j is the time spent from the onset of processing to the completion of stage j . Notice that for generality, processor j can be completed at any stage. Therefore \mathbb{T}_j is not necessarily equal to S_j . Specifically, following case II of **Fig. 4**, the TCT of item n is $\mathbb{T}_n = z_{j+1} + \dots + z_n$ but the TCT for stage n is $S_n = z_{j+1} + \dots + z_n + z_1 + \dots + z_j$, where $z_i, i \in \{1, 2, \dots, j, j + 1 \dots, n\}$ is the processing time for item i .

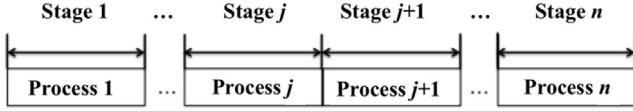
Similarly, as for the two-process standard models, we aim to investigate the joint distribution of $\mathbb{T}_{j+1}, \dots, \mathbb{T}_n$ conditional on $\mathbb{T}_1, \dots, \mathbb{T}_j$ versus the unconditional distribution function of $\mathbb{T}_{j+1}, \dots, \mathbb{T}_n$. That is

$$P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau | \mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau)$$

$$- P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau). \tag{2}$$

Now, if $\mathbb{T}_{j+1}, \dots, \mathbb{T}_n$ are finished before $\mathbb{T}_1, \dots, \mathbb{T}_j$, since each of $\mathbb{T}_1, \dots, \mathbb{T}_j$ includes the amount of $\max(\mathbb{T}_{j+1}, \dots, \mathbb{T}_n)$, the probability that $\mathbb{T}_{j+1}, \dots, \mathbb{T}_n$ are less than τ , given that each of

Case I



Case II

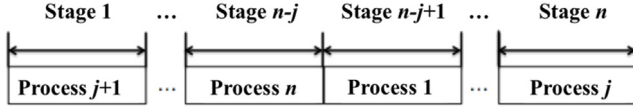


Fig. 4. Examples of possible serial arrangements for n processors.

$\mathbb{T}_1, \dots, \mathbb{T}_j$ already $< \tau$, must be 1. If $\mathbb{T}_1, \dots, \mathbb{T}_j$ are done first by time τ , the probability that $\mathbb{T}_{j+1}, \dots, \mathbb{T}_n$ also is finished by τ is greater than its marginal probability. Thus, this suggests that the TCT in a n -process serial model should be positively dependent. We will learn that this reasoning is incorrect without restriction on the processing order but correct when only one processing order is allowed.

At this point, we do not have a completely analytic proof for arbitrary n . However, we will write down the general expressions and then pick an arbitrary value of n with which to perform the pertinent computations. Theorem 9 is proved based on simulating the function derived from (2).

Theorem 9. For a standard n -process serial model,

$$P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau | \mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau) \\ - P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau)$$

can be either positive or nonpositive for $\tau > 0$.

Proof. First note that

$$P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau | \mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau) \\ = \frac{P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau, \mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau)}{P(\mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau)}.$$

We have

$$P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau, \mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau) \\ = P[\max(\mathbb{T}_{j+1}, \dots, \mathbb{T}_n, \mathbb{T}_1, \dots, \mathbb{T}_j) \leq \tau] \\ = P(z_1 + \dots + z_n \leq \tau) \\ = \int_0^\tau \dots \int_0^{\tau_3} \int_0^{\tau_2} f(\tau_n - \tau_{n-1}) \dots f(\tau_2 - \tau_1) f(\tau_1) d\tau_1 d\tau_2 \dots d\tau_n \\ = f(\tau)^{*n} * F(\tau),$$

$$\text{where } f(\tau)^{*m} = \underbrace{f(\tau) * \dots * f(\tau)}_{m \text{ times}}.$$

Now, we need to compute $P(\mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau)$. Since there is no restriction on the order of processing, it is possible to finish j processors in the j th stage and any stage after that. We observe that

$$\max(\mathbb{T}_1, \dots, \mathbb{T}_j) \in \{S_j, S_{j+1}, \dots, S_n\}.$$

We denote the probability that processors 1 to j finish in stage i as p_i , such that

$$p_i = P[\max(\mathbb{T}_1, \dots, \mathbb{T}_j) = S_i],$$

where $i \in \{j, j+1, \dots, n-1\}$ and the probability that processors 1 to j finish in stage n as q , such that

$$q = P[\max(\mathbb{T}_1, \dots, \mathbb{T}_j) = S_n].$$

It is apparent that $p_j + p_{j+1} + \dots + p_{n-1} + q = 1$. The cumulative distribution function of the TCT for the i th stage is

$$P(S_i \leq \tau) = \int_0^\tau \dots \int_0^{\tau_3} \int_0^{\tau_2} f(\tau_i - \tau_{i-1}) \dots f(\tau_2 - \tau_1) \\ \times f(\tau_1) d\tau_1 d\tau_2 \dots d\tau_i = f(\tau)^{*i} * F(\tau)$$

and for the last stage is

$$P(S_n \leq \tau) = P(t_1 + \dots + t_n \leq \tau) \\ = \int_0^\tau \dots \int_0^{\tau_3} \int_0^{\tau_2} f(\tau_n - \tau_{n-1}) \dots f(\tau_2 - \tau_1) f(\tau_1) d\tau_1 d\tau_2 \dots d\tau_n \\ = f(\tau)^{*n} * F(\tau).$$

Therefore $P(\mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau)$ can be written as

$$P(\mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau) \\ = P[\max(\mathbb{T}_1, \dots, \mathbb{T}_j) \leq \tau] \\ = p_j P(S_j \leq \tau) + \dots + p_{n-1} P(S_{n-1} \leq \tau) + q P(S_n \leq \tau) \\ = p_j [f(\tau)^{*j} * F(\tau)] + \dots + p_{n-1} [f(\tau)^{*n-1} * F(\tau)] \\ + q [f(\tau)^{*n} * F(\tau)].$$

Next, we want to compute $P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau)$. There are $n-j$ processors to be finished. They can be fully executed in the $(n-j)$ th stage and any stage after that. Therefore, we have

$$\max(\mathbb{T}_{j+1}, \dots, \mathbb{T}_n) \in \{S_{n-j}, S_{n-j+1}, \dots, S_n\}.$$

Here we denote the probability that processor $j+1$ to processor n finish in stage k as q_k , such that

$$q_k = P[\max(\mathbb{T}_{j+1}, \dots, \mathbb{T}_n) = S_k],$$

where $k \in \{n-j, n-j+1, \dots, n-1\}$. Recall that q denotes the probability that processor 1 to processor j complete in the last stage. It is equivalent to the probability that processor $j+1$ to processor n finish before stage n . Therefore, it follows that

$$q_{n-j} + q_{n-j+1} + \dots + q_{n-1} = q$$

and following the similar reasoning,

$$P[\max(\mathbb{T}_{j+1}, \dots, \mathbb{T}_n) = S_n] = p_j + p_{j+1} + \dots + p_{n-1}.$$

Combining the above two equations together, we have

$$p_j + p_{j+1} + \dots + p_{n-1} + q_{n-j} + q_{n-j+1} + \dots + q_{n-1} = 1.$$

Thus,

$$P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau) \\ = P[\max(\mathbb{T}_{j+1}, \dots, \mathbb{T}_n) \leq \tau] \\ = q_{n-j} P(S_{n-j} \leq \tau) + \dots + q_{n-1} P(S_{n-1} \leq \tau) + (p_j + p_{j+1} + \dots \\ + p_{n-1}) P(S_n \leq \tau) \\ = q_{n-j} [f(\tau)^{*n-j} * F(\tau)] + \dots + q_{n-1} [f(\tau)^{*n-1} * F(\tau)] \\ + (p_j + p_{j+1} + \dots + p_{n-1}) [f(\tau)^{*n} * F(\tau)].$$

Therefore, we arrive at the complicated expression

$$P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau | \mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau) - P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau) \\ = \frac{f(\tau)^{*n} * F(\tau)}{p_j [f(\tau)^{*j} * F(\tau)] + \dots + p_{n-1} [f(\tau)^{*n-1} * F(\tau)] + q [f(\tau)^{*n} * F(\tau)]} \\ - \{q_{n-j} [f(\tau)^{*n-j} * F(\tau)] \\ + \dots + q_{n-1} [f(\tau)^{*n-1} * F(\tau)] + (p_j + p_{j+1} + \dots + p_{n-1}) [f(\tau)^{*n} * \\ F(\tau)]\}. \quad (3)$$

We are now in a position to investigate the sign of (3). (3) takes account of all the possible permutations for the n processors. If we consider two permutations as illustrated in Fig. 4, then

$$p_{j+1} = \dots = p_{n-1} = q_{n-j+1} = \dots = q_{n-1} = 0.$$

If we can prove (3) can be both positive and nonpositive for the two permutations, the statement of this theorem will be proven for a given value of n . At this point we see that

$$\begin{aligned} (3) &= \frac{f(\tau)^{*n-1} * F(\tau)}{p_j [f(\tau)^{*j-1} * F(\tau)] + q_{n-j} [f(\tau)^{*n-1} * F(\tau)]} \\ &\quad - \{q_{n-j} [f(\tau)^{*n-j-1} * F(\tau)] \\ &\quad + p_j [f(\tau)^{*n-1} * F(\tau)]\} \\ &= R'' \left\{ 1 - p_j q_{n-j} \left[\frac{[f(\tau)^{*n-1} * F(\tau)]^{\frac{1}{2}}}{[f(\tau)^{*n-j-1} * F(\tau)]^{\frac{1}{2}} [f(\tau)^{*j-1} * F(\tau)]^{\frac{1}{2}}} \right]^2 \right. \\ &\quad \left. + p_j q_{n-j} \left\{ [f(\tau)^{*j-1} * F(\tau)]^{\frac{1}{2}} - [f(\tau)^{*n-j-1} * F(\tau)]^{\frac{1}{2}} \right\}^2 \right. \\ &\quad \left. - p_j [f(\tau)^{*j-1} * F(\tau)] - q_{n-j} [f(\tau)^{*n-j-1} * F(\tau)] \right\}, \quad (4) \end{aligned}$$

where $R'' = \frac{f(\tau)^{*n-1} * F(\tau)}{p_j [f(\tau)^{*j-1} * F(\tau)] + q_{n-j} [f(\tau)^{*n-1} * F(\tau)]}$ and $p_j + q_{n-j} = 1$. If $f(\tau)^{*j-1} * F(\tau) = f(\tau)^{*n-j-1} * F(\tau)$, that is $j = \frac{n}{2}$, and then (4) is reduced to

$$R'' \left\{ 1 - p_j q_{n-j} \left\{ [f(\tau)^{*n-1} * F(\tau)]^{\frac{1}{2}} - \frac{f(\tau)^{*j-1} * F(\tau)}{[f(\tau)^{*n-1} * F(\tau)]^{\frac{1}{2}}} \right\}^2 - [f(\tau)^{*j-1} * F(\tau)] \right\}.$$

Since $p_j q_{n-j} \leq \frac{1}{4}$, the above expression is

$$\begin{aligned} &\geq R'' \left\{ 1 - \frac{1}{4} \left\{ [f(\tau)^{*n-1} * F(\tau)]^{\frac{1}{2}} - \frac{f(\tau)^{*j-1} * F(\tau)}{[f(\tau)^{*n-1} * F(\tau)]^{\frac{1}{2}}} \right\}^2 \right. \\ &\quad \left. - [f(\tau)^{*j-1} * F(\tau)] \right\}. \quad (5) \end{aligned}$$

Note that if $p_j = \frac{1}{2}$, the above \geq reduces to $=$. Now, it is requisite to investigate the sign of (5). If it can be both positive and nonpositive, it indicates that (2) can be positive and nonpositive. We have this ordering:

$$\begin{aligned} 0 &< f(\tau)^{*n-1} * F(\tau) \leq f(\tau)^{*j-1} * F(\tau) \leq 1 \\ &\leq \frac{f(\tau)^{*j-1} * F(\tau)}{[f(\tau)^{*n-1} * F(\tau)]^{\frac{1}{2}}}, \end{aligned}$$

as $\frac{f(\tau)^{*j-1} * F(\tau)}{[f(\tau)^{*n-1} * F(\tau)]^{\frac{1}{2}}} = \left[\frac{f(\tau)^{*j-1} * F(\tau) f(\tau)^{*j-1} * F(\tau)}{f(\tau)^{*n-1} * F(\tau)} \right]^{\frac{1}{2}}$
 $= \left[\frac{f(\tau)^{*j-1} * F(\tau) f(\tau)^{*n-j-1} * F(\tau)}{f(\tau)^{*n-1} * F(\tau)} \right]^{\frac{1}{2}} = \left[\frac{P(S_j \leq \tau) P(S_{n-j} \leq \tau)}{P(S_n \leq \tau)} \right]^{\frac{1}{2}} = \left[\frac{P(\max\{S_j, S_{n-j}\} \leq \tau)}{P(S_n \leq \tau)} \right]^{\frac{1}{2}} \geq 1$. The sign of (5) was computed using the simulation-based method the same as that was used for $n = 2$ following these steps:

Step 1: generate a random number $\alpha \sim Uniform[0, 1]$, where α represents $f(\tau)^{*n-1} * F(\tau)$.

Step 2: generate a random number $\beta \sim Uniform[\alpha, 1]$, where β represents $f(\tau)^{*j-1} * F(\tau)$.

Step 3: if $\frac{\beta^2}{\alpha} \geq 1$, then compute if the part after R'' for (5) > 0 .

It was found that the probability of the part after R'' for (5) > 0 given $\frac{\beta^2}{\alpha} \geq 1$ is 62%. The simulation result indicates that (5) can be either positive or nonpositive. □

We proceed constructing several examples that help to clarify Theorem 9. We assume that there are three processors in a standard serial model, whose processing times are i.i.d and labeled as:

$$Z_1, Z_2, Z_3.$$

Of course, one can construct examples that have more than three processors in the models. Here we only discuss standard three-process serial models. Adapting the above expressions, and without loss of generality, one is interested in the signs of the two functions below:

$$\begin{aligned} &P(\mathbb{T}_1 \leq \tau | \mathbb{T}_2 \leq \tau, \mathbb{T}_3 \leq \tau) - P(\mathbb{T}_1 \leq \tau), \\ &P(\mathbb{T}_1 \leq \tau, \mathbb{T}_2 \leq \tau | \mathbb{T}_3 \leq \tau) - P(\mathbb{T}_1 \leq \tau, \mathbb{T}_2 \leq \tau). \end{aligned}$$

We will only discuss the behavior of first of the two above functions. The other function can be examined in an analogous fashion.

According to the earlier definition of \mathbb{T} and \mathbb{S} , we see that

$$\begin{aligned} \max(\mathbb{T}_2, \mathbb{T}_3) &\in \{\mathbb{S}_2, \mathbb{S}_3\}, \\ \mathbb{T}_1 &\in \{\mathbb{S}_1, \mathbb{S}_2, \mathbb{S}_3\}. \end{aligned}$$

We subsequently denote

$$\begin{aligned} P[\max(\mathbb{T}_2, \mathbb{T}_3) = \mathbb{S}_2] &= p_2, \\ P[\mathbb{T}_1 = \mathbb{S}_1] &= q_1, \\ P[\mathbb{T}_1 = \mathbb{S}_2] &= q_2. \end{aligned}$$

Then apparently, we have

$$\begin{aligned} P[\max(\mathbb{T}_2, \mathbb{T}_3) = \mathbb{S}_3] &= q_1 + q_2, \\ P[\mathbb{T}_1 = \mathbb{S}_3] &= p_2, \end{aligned}$$

and

$$p_2 + q_1 + q_2 = 1.$$

Consequently,

$$\begin{aligned} &P(\mathbb{T}_2 \leq \tau, \mathbb{T}_3 \leq \tau) \\ &= P(\max(\mathbb{T}_2, \mathbb{T}_3) \leq \tau) \\ &= p_2 P(\mathbb{S}_2 \leq \tau) + (q_1 + q_2) P(\mathbb{S}_3 \leq \tau) \\ &= p_2 f(\tau) * F(\tau) + (q_1 + q_2) f(\tau) * f(\tau) * F(\tau), \\ &P(\mathbb{T}_1 \leq \tau) \\ &= q_1 P(\mathbb{S}_1 \leq \tau) + q_2 P(\mathbb{S}_2 \leq \tau) + p_2 P(\mathbb{S}_3 \leq \tau) \\ &= q_1 F(\tau) + q_2 f(\tau) * F(\tau) + p_2 f(\tau) * f(\tau) * F(\tau). \end{aligned}$$

Therefore,

$$\begin{aligned} &P(\mathbb{T}_1 \leq \tau | \mathbb{T}_2 \leq \tau, \mathbb{T}_3 \leq \tau) - P(\mathbb{T}_1 \leq \tau) \\ &= \frac{P(\mathbb{T}_1 \leq \tau, \mathbb{T}_2 \leq \tau, \mathbb{T}_3 \leq \tau)}{P(\mathbb{T}_2 \leq \tau, \mathbb{T}_3 \leq \tau)} - P(\mathbb{T}_1 \leq \tau) \\ &= \frac{f(\tau) * f(\tau) * F(\tau)}{p_2 f(\tau) * F(\tau) + (q_1 + q_2) f(\tau) * f(\tau) * F(\tau)} \\ &\quad - (q_1 F(\tau) + q_2 f(\tau) * F(\tau) + p_2 f(\tau) * f(\tau) * F(\tau)). \quad (6) \end{aligned}$$

Let us consider some specific often employed, distributions.

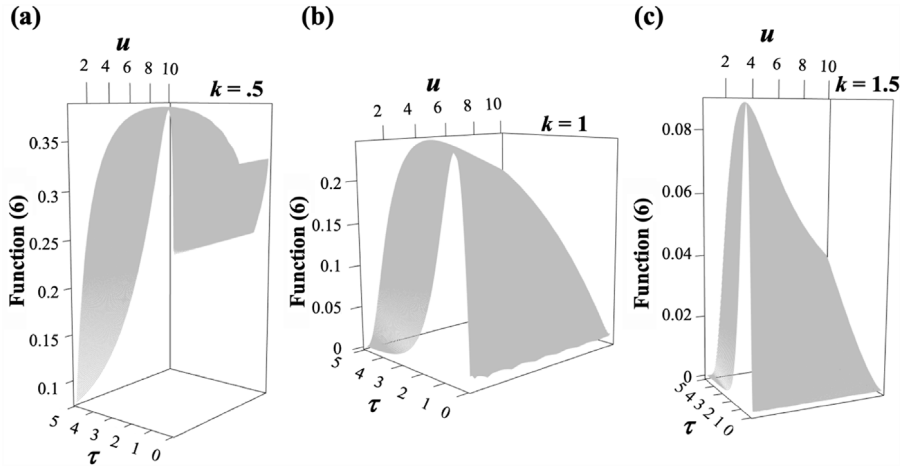


Fig. 5. The plots of function (6) for (a) $k = .5$, (b) $k = 1$, and (c) $k = 1.5$. Note that τ has the arbitrary unit.

Weibull distributions. Let z_1, z_2, z_3 be i.i.d and follow the Weibull distribution with the density function

$$f(\tau) = ku(u\tau)^{k-1} \exp[-(u\tau)^k],$$

where the parameters $k, u > 0$. We used simulation methods to compute the values of (6) by selecting $p_2 = q_1 = q_2 = \frac{1}{3}$. Here we present plots for (6) by varying the values of τ and u (Fig. 5). We allowed u to vary from .5 to 10 and τ to vary from 0.01 to 5. Fig. 5(a) fixes $k = .5$, Fig. 5(b) fixes $k = 1$, where the Weibull distributions reduce to exponential distributions, and Fig. 5(c) fixes $k = 1.5$. The three plots for (6) are non-negative.

Next, we explore the uniform distribution. Although perhaps, unrealistic in fact, it has often been employed to represent the so-called base time distribution, namely the duration which captures the additional unmodeled psychological processes such as early sensory and late motor stages.

Uniform distributions. Let z_1, z_2, z_3 be i.i.d and follow the uniform distribution

$$z_1, z_2, z_3 \sim \text{Uniform}(0, v),$$

where $v > 0$. The corresponding distribution function is

$$F(\tau) = \begin{cases} \frac{\tau}{v}, & \text{if } 0 \leq \tau \leq v \\ 1, & \text{otherwise.} \end{cases}$$

Also, we have

$$f(\tau) * F(\tau) = \begin{cases} \int_0^\tau \frac{\tau_2}{v^2} d\tau_2 = \frac{\tau^2}{2v^2}, & \text{if } 0 \leq \tau < v, \\ \int_v^\tau \frac{2v - \tau_2}{v^2} d\tau_2 + \frac{1}{2} = \frac{2\tau}{v} - \frac{\tau^2}{2v^2} - 1, & \text{if } v \leq \tau < 2v, \\ 1, & \text{if } 2v \leq \tau. \end{cases}$$

$$f(\tau) * f(\tau) * F(\tau) = \begin{cases} \int_0^\tau \frac{\tau_3^2}{2v^3} d\tau_3 = \frac{\tau^3}{6v^3}, & \text{if } 0 \leq \tau < v, \\ \int_v^\tau \left(-\frac{\tau_3^2}{v^3} + \frac{3\tau_3}{v^2} - \frac{3}{2v} \right) d\tau_3 + \frac{1}{6} = -\frac{\tau^3}{3v^3} + \frac{3\tau^2}{2v^2} - \frac{3\tau}{2v} + \frac{1}{2}, & \text{if } v \leq \tau < 2v, \\ \int_{2v}^\tau \left(\frac{\tau_3^2}{2v^3} - \frac{3\tau_3}{v^2} + \frac{9}{2v} \right) d\tau_3 + \frac{5}{6} = \frac{\tau^3}{6v^3} - \frac{3\tau^2}{2v^2} + \frac{9\tau}{2v} - \frac{7}{2}, & \text{if } 2v \leq \tau < 3v, \\ 1, & \text{if } 3v \leq \tau. \end{cases}$$

For $3v \leq \tau$,

$$\begin{aligned} &P(\mathbb{T}_1 \leq \tau | \mathbb{T}_2 \leq \tau, \mathbb{T}_3 \leq \tau) - P(\mathbb{T}_1 \leq \tau) \\ &= \frac{f(\tau) * f(\tau) * F(\tau)}{p_2 f(\tau) * F(\tau) + (q_1 + q_2) f(\tau) * f(\tau) * F(\tau) - (q_1 F(\tau) + q_2 f(\tau) * F(\tau) + p_2 f(\tau) * f(\tau) * F(\tau))} \\ &= \frac{1}{p_2 + (q_1 + q_2)} - (q_1 + q_2 + p_2) \\ &= 0; \end{aligned}$$

for $2v \leq \tau < 3v$,

$$\begin{aligned} &P(\mathbb{T}_1 \leq \tau | \mathbb{T}_2 \leq \tau, \mathbb{T}_3 \leq \tau) - P(\mathbb{T}_1 \leq \tau) \\ &= \frac{f(\tau) * f(\tau) * F(\tau)}{p_2 f(\tau) * F(\tau) + (q_1 + q_2) f(\tau) * f(\tau) * F(\tau) - (q_1 F(\tau) + q_2 f(\tau) * F(\tau) + p_2 f(\tau) * f(\tau) * F(\tau))} \\ &= \frac{f(\tau) * f(\tau) * F(\tau)}{p_2 + (q_1 + q_2) f(\tau) * f(\tau) * F(\tau) - (q_1 + q_2 + p_2 f(\tau) * f(\tau) * F(\tau))} \\ &= \frac{-p_2(1 - p_2)}{p_2 + (q_1 + q_2) f(\tau) * f(\tau) * F(\tau)} \\ &\quad \times [(f(\tau) * f(\tau) * F(\tau))^2 - 2f(\tau) * f(\tau) * F(\tau) + 1] \\ &\leq 0; \end{aligned}$$

for $v \leq \tau < 2v$ and $0 \leq \tau < v$, we ran simulations to investigate the behavior of (6) and the result is plotted in Fig. 6. Again, we fixed $p_2 = q_1 = q_2 = \frac{1}{3}$ and $v = 2$. By observing the plot, we conclude that (6) can be either positive or nonpositive for uniformly distributed processing times.

It is now a good place to demonstrate, for any n , that if only one order of processing is allowed, the dependence of total completion times is always nonnegative. This too, generalizes our earlier results for $n = 2$.

Corollary 10. For a standard n -process serial model, $P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau | \mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau) - P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau)$ is non-negative for $\tau > 0$ if only one permutation is allowed.

Proof. According to the derivation in Theorem 9, we have

$$\begin{aligned} &P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau | \mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau) - P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau) \\ &= \frac{f(\tau)^{*(n-1)} * F(\tau)}{p_j [f(\tau)^{*(j-1)} * F(\tau)] + \dots + p_{n-1} [f(\tau)^{*(n-2)} * F(\tau)] + q [f(\tau)^{*(n-1)} * F(\tau)]} \\ &\quad - \{q_{n-j} [f(\tau)^{*(n-j-1)} * F(\tau)] + \dots + q_{n-1} [f(\tau)^{*(n-2)} * F(\tau)] \\ &\quad + (p_j + p_{j+1} + \dots + p_{n-1}) [f(\tau)^{*(n-1)} * F(\tau)]\}. \end{aligned}$$

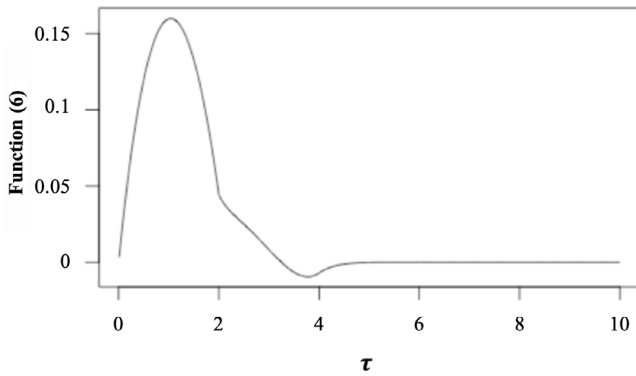


Fig. 6. The plot of function (6) for uniformly distributed processing times. Note that τ has the arbitrary unit.

If one permutation is allowed, one and only one element in the set

$$\{p_j, p_{j+1}, \dots, p_{n-1}, q_{n-j}, q_{n-j+1}, \dots, q_{n-1}\}$$

is 1 and the other elements are 0s. So, the above equation can be written as

$$\begin{aligned} &= R \{f(\tau)^{* (n-1)} * F(\tau) \\ &\quad - \{q_{n-j} [f(\tau)^{* (n-j-1)} * F(\tau)] + \dots + q_{n-1} [f(\tau)^{* (n-2)} * F(\tau)] \\ &\quad + (p_j + p_{j+1} + \dots + p_{n-1}) [f(\tau)^{* (n-1)} * F(\tau)]\} \\ &\quad \times \{p_j [f(\tau)^{* (j-1)} * F(\tau)] + \dots \\ &\quad + p_{n-1} [f(\tau)^{* (n-2)} * F(\tau)] + q [f(\tau)^{* (n-1)} * F(\tau)]\} \\ &= R \{f(\tau)^{* (n-1)} * F(\tau) - f(\tau)^{* (l-1)} * F(\tau) [f(\tau)^{* (n-1)} * F(\tau)]\} \\ &\geq 0 \end{aligned}$$

for $\tau > 0$, where

$$R = \frac{1}{p_j [f(\tau)^{* (j-1)} * F(\tau)] + \dots + p_{n-1} [f(\tau)^{* (n-2)} * F(\tau)] + q [f(\tau)^{* (n-1)} * F(\tau)]}$$

and

$$l \in \{\min(j, n-j), \dots, n-1\}. \quad \square$$

Finally, we also can show that the coarser statistic, the covariance, between any two, of the subprocesses, is always positive.

Lemma 11. For a standard n -process serial model, again assuming a single processing order,

$$\text{Cov}(\mathbb{T}_j, \mathbb{T}_l) = \sum_{i=k}^n \sum_{z=k}^n \sum_{k=1}^n m_i w_z \text{Var}(T_k), \quad 1 \leq j, l \leq n, \text{ and } l \neq j.$$

Here $\mathbb{T}_j = \sum_{i=1}^n m_i \mathbb{S}_i$, where m_i is the probability for $\mathbb{T}_j = \mathbb{S}_i$ and $\sum_{i=1}^n m_i = 1$. $\mathbb{T}_l = \sum_{z=1}^n w_z \mathbb{S}_z$, where w_z is the probability for $\mathbb{T}_l = \mathbb{S}_z$ and $\sum_{z=1}^n w_z = 1$.

Proof.

$$\begin{aligned} \text{Cov}(\mathbb{T}_j, \mathbb{T}_l) &= \text{Cov}\left(\sum_{i=1}^n m_i \mathbb{S}_i, \sum_{z=1}^n w_z \mathbb{S}_z\right) \\ &= \text{Cov}\left(\sum_{i=1}^n m_i \sum_{k=1}^i T_k, \sum_{z=1}^n w_z \sum_{k=1}^z T_k\right) \end{aligned}$$

$$\begin{aligned} &= \text{Cov}\left(\sum_{i=1}^n \sum_{k=1}^i m_i T_k, \sum_{z=1}^n \sum_{k=1}^z w_z T_k\right) \\ &= \sum_{i=k}^n \sum_{z=k}^n \sum_{k=1}^n m_i w_z \text{Var}(T_k). \quad \square \end{aligned}$$

In standard multiple-process parallel models, by definition, the TCT for a processor is the processing time for that channel. That implies that the TCTs for processors $\mathbb{T}_1, \dots, \mathbb{T}_n$ are identically and independently distributed. Similarly, as for the standard n -process serial models, we state the behavior of (2) for the standard n -process parallel models, which we know from the prior knowledge, to possess no dependence. We include the statement of these tautologies to contrast the distinct predictions of the standard serial vs. the standard parallel models.

Theorem 12. For a standard n -process parallel model, $P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau | \mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau) - P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau) = 0$ for $\tau > 0$.

Proof. For a standard n -process parallel model,

$$\begin{aligned} &P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau | \mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau) \\ &\quad - P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau) \\ &= \frac{P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau, \mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau)}{P(\mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau)} \\ &\quad - P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau) \\ &= \frac{P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau) P(\mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau)}{P(\mathbb{T}_1 \leq \tau, \dots, \mathbb{T}_j \leq \tau)} \\ &\quad - P(\mathbb{T}_{j+1} \leq \tau, \dots, \mathbb{T}_n \leq \tau) \\ &= 0. \quad \square \end{aligned}$$

And, of course, the covariance is destined to be 0 as well.

Lemma 13. For a standard n -process parallel model,

$$\text{Cov}(\mathbb{T}_j, \mathbb{T}_l) = 0,$$

where $1 \leq j, l \leq n$, and $l \neq j$.

Proof. It is apparent. \square

Our investigation of TCTs for general n shows that dependence of these in the case of standard serial models can be either positive or nonpositive. The computations of special cases strongly suggest that positive dependencies are much easier to come by, perhaps because of our earlier expressed intuitions in the section for the standard two-process models.

Furthermore, the above derivations suggest that the behavior of standard serial and standard parallel models differ substantively. Function (2) = 0 looks difficult for a distribution for serial class of models to satisfy for all $\tau > 0$. However, it is a functional equation that appears to be quite challenging to solve. On the other hand, the case for a single processing order is concrete and clear: Function (2) has to be 0 for a standard parallel model yet cannot be for the standard serial model with a single order. The strategic issue of model mimicry will be revisited in a subsequent section.

3.2. Dependence of intercompletion times, general n

We now proceed to investigate the general statistic associated with the actual processing times of standard serial models, that is, the ICTs. Without loss of generalization, we can assume that $z_1 \leq z_2 \leq \dots \leq z_n$. Recall that in standard n -process serial

models, the ICTs are independent. Therefore, we only investigate the dependence of ICTs for standard n -process parallel models.

The standard n -process parallel model can be decomposed into n stages (Fig. 1(b) can be viewed as a standard n -process parallel model). We have the ICTs for the standard n -process parallel models:

$$\begin{aligned} T_1 &= \mathbb{T}_1 = z_1, \\ T_2 &= \mathbb{T}_2 - \mathbb{T}_1 = z_2 - z_1, \\ &\dots, \\ T_n &= \mathbb{T}_n - \mathbb{T}_{n-1} = z_n - z_{n-1}. \end{aligned}$$

We investigate the behavior of the survival function of the canonical sum of ICTs $T_{j+1} + \dots + T_{n'}$, $0 < j + 1 \leq n' \leq n$, conditional on the completion of the earlier stages: $P(T_{j+1} + \dots + T_{n'} > t | z_{n'} > z_j)$. Without loss of generality, we assume that processor 1 terminates first, processor 2 terminates second, ..., and processor n terminates in the last. Lemma 14 shows that the ICTs of standard parallel models tend to increase as a function of the previous ICTs in a strong distributional sense if the hazard function is constant or decreasing.

Lemma 14. For a standard n -process parallel model, if the hazard function h is non-increasing, then $P(T_{j+1} + \dots + T_{n'} > t | z_{n'} > z_j)$ is non-decreasing as z_j is increased.

Proof. We have

$$\begin{aligned} &P(T_{j+1} + \dots + T_{n'} > t | z_{n'} > z_j) \\ &= P(T_1 + \dots + T_j + T_{j+1} + \dots + T_{n'} > t \\ &\quad + T_1 + \dots + T_j | z_{n'} > z_j) \\ &= P(z_{n'} > t + z_j | z_{n'} > z_j) \\ &= \frac{S(z_j + t)}{S(z_j)}. \end{aligned}$$

To examine the behavior of this function as z_j changes, one can take the derivative

$$\begin{aligned} \frac{d}{dz_j} \frac{S(z_j + t)}{S(z_j)} &= \frac{-S(z_j)f(z_j + t) + S(z_j + t)f(z_j)}{S^2(z_j)} \\ &= \frac{S(z_j)S(z_j + t)}{S^2(z_j)} \left[\frac{f(z_j)}{S(z_j)} - \frac{f(z_j + t)}{S(z_j + t)} \right] \\ &= \frac{S(z_j)S(z_j + t)}{S^2(z_j)} [h(z_j) - h(z_j + t)]. \end{aligned}$$

If the hazard function is non-increasing, then $\frac{d}{dz_j} \frac{S(z_j + t)}{S(z_j)} \geq 0$. Consequently, $P(T_{j+1} + \dots + T_{n'} > t | z_{n'} > z_j)$ is non-decreasing as z_j is increased. \square

Thus, happily we again find that, as for $n = 2$, a constant or decreasing hazard function forces non-decreasing sums of ICTs.

Next, as before with $n = 2$, we move on to deliberate on what characterizes the successive ICTs for standard parallel models in the general case of arbitrary n . We can write the survival functions for the n -process parallel models at each stage conditioned on the completion of the earlier stage(s) in Table 1:

Let us denote the ratio of the hazard functions

$$\alpha(z_{j-1} + t, z_j + t) = \frac{h(z_j + t)}{h(z_{j-1} + t)}.$$

If the survival function from stage j to stage $j + 1$ for every j is increasing, this trend is then consistent with the empirical finding in free recall tasks that the time interval between two successive

Table 1
Survival functions for the standard n -process parallel models at each stage.

Stage	Survival function
Stage 1	$S^n(t)$
Stage 2	$\left[\frac{S(z_1 + t)}{S(z_1)} \right]^{n-1}$
Stage 3	$\left[\frac{S(z_2 + t)}{S(z_2)} \right]^{n-2}$
...	...
Stage n	$\frac{S(z_{n-1} + t)}{S(z_{n-1})}$

responses is increasing as stage number is increasing. Theorem 15 provides under what exact condition the survival function keeps increasing from stage j to stage $j + 1$. A generalized form of the result for $n = 2$ again finds that the above ratio of hazard functions controls the behavior of the sequence of ICTs. Corollary 16 states that standard n -process parallel models with concave or linear $H(t)$ result in the increasing survival functions across stages.

Theorem 15. For a standard n -process parallel model, the survivor function from the j th stage to the $(j + 1)$ th stage is increasing if $\alpha(z_{j-1} + t, z_j + t) < \frac{n-j+1}{n-j}$ and is non-increasing otherwise.

Proof. Consider stage j vs. stage $j + 1$:

$$\begin{aligned} &\left[\frac{S(t + z_{j-1})}{S(z_{j-1})} \right]^{n-j+1} - \left[\frac{S(t + z_j)}{S(z_j)} \right]^{n-j} \\ &= \frac{\exp[-(n-j+1)H(t + z_{j-1})]}{\exp[-(n-j+1)H(z_{j-1})]} - \frac{\exp[-(n-j)H(t + z_j)]}{\exp[-(n-j)H(z_j)]}. \end{aligned}$$

The sign of the above equation is the same as

$$-(n-j+1)H(t + z_{j-1}) - (n-j)H(z_j) + (n-j)H(t + z_j) + (n-j+1)H(z_{j-1}), \tag{7}$$

which is equivalent to

$$\begin{aligned} &-(n-j+1) \int_0^t h(z_{j-1} + t) dt + (n-j) \int_0^t h(z_j + t) dt \\ &= \int_0^t -(n-j+1)h(z_{j-1} + t) + (n-j)\alpha(z_{j-1} + t, z_j + t) \\ &\quad \times h(z_{j-1} + t) dt \\ &= \int_0^t h(z_{j-1} + t) [-(n-j+1) + (n-j)\alpha(z_{j-1} + t, z_j + t)] dt \end{aligned}$$

$$\begin{cases} \geq 0, & \text{if } \alpha(z_{j-1} + t, z_j + t) \geq \frac{n-j+1}{n-j} \\ < 0, & \text{if } \alpha(z_{j-1} + t, z_j + t) < \frac{n-j+1}{n-j}. \quad \square \end{cases} \tag{8}$$

Corollary 16. For a standard n -process parallel model, (i) if the cumulative hazard function $H(t)$ is concave or linear, then

$$\left[\frac{S(t + z_{j-1})}{S(z_{j-1})} \right]^{n-j+1} - \left[\frac{S(t + z_j)}{S(z_j)} \right]^{n-j} < 0;$$

(ii) if $H(t)$ is convex, then the sign of $\left[\frac{S(t+z_{j-1})}{S(z_{j-1})} \right]^{n-j+1} - \left[\frac{S(t+z_j)}{S(z_j)} \right]^{n-j}$ is uncertain.

Proof. (i) If the cumulative hazard function $H(t)$ is concave or linear, then the hazard function $h(t)$ is a decreasing or constant. Consequently, $\alpha(z_{j-1} + t, z_j + t) < 1$ or $= 1$. According to

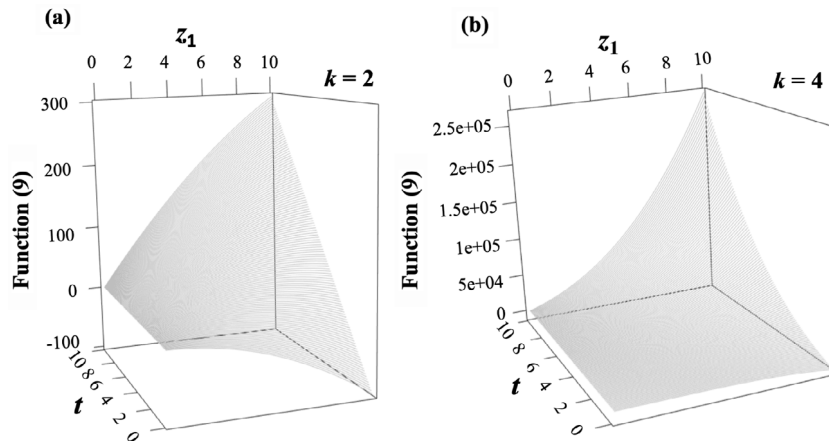


Fig. 7. Plots of function (9) for (a) $k = 2$ and (b) $k = 4$. Note that t and z_1 have the arbitrary and the same unit.

Theorem 15. $\left[\frac{s(t+z_{j-1})}{s(z_{j-1})}\right]^{n-j+1} - \left[\frac{s(t+z_j)}{s(z_j)}\right]^{n-j} < 0$ is obtained. (ii) If $H(t)$ is convex, the hazard function $h(t)$ is increasing. Then we have $\alpha(z_{j-1} + t, z_j + t) = \frac{h(z_j+t)}{h(z_{j-1}+t)} > 1$. It is uncertain if $\alpha(z_{j-1} + t, z_j + t) \geq \frac{n-j+1}{n-j}$ or $\alpha(z_{j-1} + t, z_j + t) < \frac{n-j+1}{n-j}$. So, the sign of $\left[\frac{s(t+z_{j-1})}{s(z_{j-1})}\right]^{n-j+1} - \left[\frac{s(t+z_j)}{s(z_j)}\right]^{n-j}$ is uncertain. \square

Here we construct examples to further illustrate the behavior of $\left[\frac{s(t+z_{j-1})}{s(z_{j-1})}\right]^{n-j+1} - \left[\frac{s(t+z_j)}{s(z_j)}\right]^{n-j}$. As before, we learn that standard parallel models are characterized by a vital inclination toward ever-increasing intervals between successive completions, that is $\alpha(z_{j-1} + t, z_j + t) \geq \frac{n-j+1}{n-j}$. But that tendency can be overridden by extremely increasing hazard functions. These findings are illustrated in the subsequent examples and Figs. 7–9.

To construct examples for $n > 2$, we assume that there are three processors in standard parallel models, whose processing times are i.i.d and labeled as:

$$z_1, z_2, z_3.$$

According to Theorem 15, the survival function from stage j to stage $j + 1$ is increasing if $\alpha(z_{j-1} + t, z_j + t) < \frac{n-j+1}{n-j}$. For $n = 3$, that is

$$\alpha(t, z_1 + t) = \frac{h(z_1 + t)}{h(t)} < \frac{3 - 1 + 1}{3 - 1} = \frac{3}{2}, \text{ if } j = 1,$$

$$\alpha(t + z_1, t + z_2) = \frac{h(z_2 + t)}{h(z_1 + t)} < \frac{3 - 2 + 1}{3 - 2} = 2, \text{ if } j = 2.$$

For $j = 1$ and 2, (7) is equivalent to

$$-3H(t) - 2H(z_1) + 2H(t + z_1), \tag{9}$$

and

$$-2H(t + z_1) - H(z_2) + H(t + z_2) + 2H(z_1), \tag{10}$$

respectively.

Weibull distributions. Let

$$z_1, z_2, z_3 \sim \text{Weibull}(k, u),$$

where $k, u > 0$. The corresponding cumulative hazard function and the hazard function are

$$H(t) = u(ut)^{k-1}t$$

and

$$h(t) = uk(ut)^{k-1}.$$

If $k = 1$, then Weibull distributions reduce to exponential distributions:

$$z_1, z_2, z_3 \sim \text{Exp}(u).$$

The cumulative hazard function for the exponential distribution is linear:

$$H(t) = ut.$$

The hazard function is a constant: $h(t) = u$. It is apparent

$$\alpha(t, z_1 + t) = \frac{h(z_1 + t)}{h(t)} = 1 < \frac{3}{2}, \text{ if } j = 1,$$

$$\alpha(t + z_1, t + z_2) = \frac{h(z_2 + t)}{h(z_1 + t)} = 1 < 2, \text{ if } j = 2.$$

Therefore, the survival function for exponentially distributed processors is increasing from the first stage to the second stage to the third stage, as expected.

If $k < 1$, then the cumulative hazard function is concave and the hazard function is increasing. Hence, the survival is also increasing from stage 1 to stage 2 to stage 3 as

$$\alpha(t, z_1 + t) = \frac{h(z_1 + t)}{h(t)} < 1 < \frac{3}{2}, \text{ if } j = 1,$$

$$\alpha(t + z_1, t + z_2) = \frac{h(z_2 + t)}{h(z_1 + t)} < 1 < 2, \text{ if } j = 2.$$

If $k > 1$, then the cumulative hazard function is convex. We ran simulations to investigate the dynamics of the survival function from stage 1 to stage 2 to stage 3. We present 3d plots (Fig. 7) of (9) by varying the values of t and z_1 from 0 to 10 and fixing $u = 1$. Fig. 7(a) fixes $k = 2$ and Fig. 7(b) fixes $k = 4$. We also present 3d plots (Fig. 8) of (10) by varying the values of t and z_2 from 0 to 10 and fixing $z_1 = 1$ and $u = 1$. Fig. 8(a) fixes $k = 2$ and Fig. 8(b) fixes $k = 4$. We observe both (9) and (10) can be negative and nonnegative. We conclude that for $k > 1$, the survival function from stage 1 to stage 2 to stage 3 is neither always increasing nor always non-increasing.

Uniform distributions. Let

$$z_1, z_2, z_3 \sim \text{Uniform}(0, v),$$

where $v > 0$. The corresponding cumulative hazard function ($0 \leq t \leq v$) is convex:

$$H(t) = -\ln(v - t) + \ln v$$

and the hazard function is

$$h(t) = \frac{1}{v - t}.$$

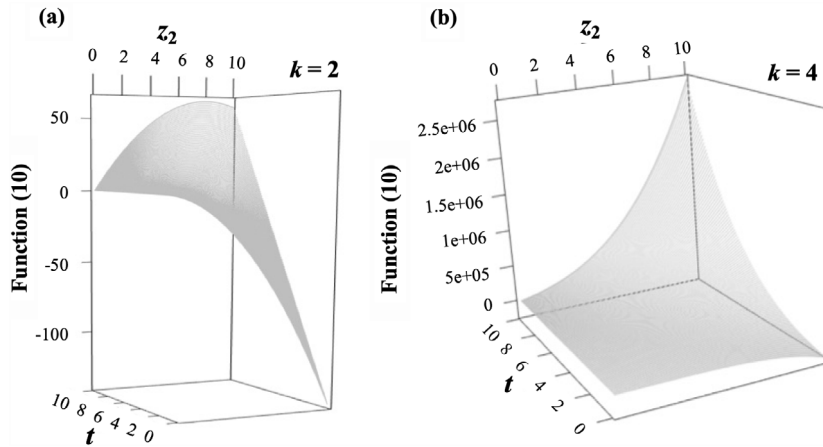


Fig. 8. Plots of function (10) for (a) $k = 2$ and (b) $k = 4$. Note that t and z_2 have the arbitrary and the same unit.

We have

$$\alpha(t, z_1 + t) = \frac{h(z_1 + t)}{h(t)} = \frac{v - t}{v - z_1 - t} \geq 1, \text{ if } j = 1,$$

$$\alpha(t + z_1, t + z_2) = \frac{h(z_2 + t)}{h(z_1 + t)} = \frac{v - z_1 - t}{v - z_2 - t} \geq 1, \text{ if } j = 2.$$

We ran simulations to investigate the dynamics of the survival from stage 1 to stage 2 to stage 3.

We present a 3d plot (Fig. 9(a)) of function (9) by varying the values of t and z_1 from 0 to 1 and fixing $v = 2$. We also present a 3d plot (Fig. 9(b)) of function (10) by varying the values of t and z_2 from 0 to 1 and fixing $z_1 = .5$ and $v = 2$. We observe both (9) and (10) can be negative and nonnegative. We conclude that for uniformly distributed processing times, the survival function from stage 1 to stage 2 to stage 3 is neither always increasing nor always non-increasing.

Analogously to the situation with TCTs, we find that in general ICTs for standard parallel models do not tend to follow the behavior of those for standard serial models which are i.i.d. However, we will proceed to discuss the pivotal issue of mimicry directly.

3.3. Can standard serial and standard parallel models ever mimic one another?

Our very first papers on serial vs. parallel processing uncovered serious issues of mathematical equivalence of these classes in popular experimental paradigms (e.g., Townsend, 1969, 1971, 1972). Later work expanded the classes of distributions subject to mimicry problems (e.g., Townsend, 1976a) as well as pointing to experimental situations and model depictions that might be capable of distinguishing them (e.g., Townsend, 1976b). Yet, those results do not answer the question of model mimicry within the special constraints of standard serial vs. standard parallel models.

In this realm, the mimicry concern devolves into the inquiry of whether a standard serial model and its assumption of i.i.d ICTs can simultaneously be not only parallel (i.e., equivalent to some arbitrary parallel model) but also produce i.i.d TCTs. The opposite way to pose the question is whether a standard parallel model can be equivalent not only to some arbitrary serial model but to one with i.i.d ICTs. But, observe that this question is actually symmetric because it simply asks whether there is an intersection of the two classes of models that includes the conjunction of i.i.d ICTs and i.i.d TCTs.

We first inspect function (3) in this paper or its special case for $n = 2$ in Zhang et al. (2018) and view either one as a functional equation when the potential inequality is set identically equal to 0. Subjectively, and as intimated earlier, it does not appear

that any continuous family of probability densities can satisfy this equation for all times τ . However, we currently have no proof of that.

Hence, we move on to Eq. (8) and check if $\alpha(z_{j-1} + t, z_j + t) \equiv \frac{n-j+1}{n-j}$ for all the values of $z_{j-1} > 0, z_j > 0, t > 0$. We recount the definition of α here that $\alpha(z_{j-1} + t, z_j + t) = \frac{h(z_j+t)}{h(z_{j-1}+t)}$. If $\alpha(z_{j-1} + t, z_j + t) \equiv \frac{n-j+1}{n-j}$ holds, it indicates standard parallel models can produce i.i.d ICTs, which is a fundamental property of standard serial models. A straightforward proof that no standard parallel model can yield this identity follows.

Lemma 17. For a standard n -process parallel model, $\alpha(z_{j-1} + t, z_j + t) \equiv \frac{n-j+1}{n-j}$ cannot hold for all the values of $z_{j-1} > 0, z_j > 0, t > 0$.

Proof. Assuming the above equation holds, then we shall have

$$\frac{h(z_j + t + (z_j - z_{j-1}))}{h(z_{j-1} + t + (z_j - z_{j-1}))} = \frac{n - j + 1}{n - j}$$

$$\frac{h(2z_j + t - z_{j-1})}{h(z_j + t)} = \frac{n - j + 1}{n - j}$$

$$\frac{h(2z_j + t - z_{j-1})}{\frac{n-j+1}{n-j} h(z_{j-1} + t)} = \frac{n - j + 1}{n - j}$$

$$\frac{h(2z_j + t - z_{j-1})}{h(z_{j-1} + t)} = \left(\frac{n - j + 1}{n - j}\right)^2$$

$$\frac{h((2z_j - z_{j-1}) + t)}{h(z_{j-1} + t)} = \left(\frac{n - j + 1}{n - j}\right)^2$$

Contradiction! So $\alpha(z_{j-1} + t, z_j + t) \equiv \frac{n-j+1}{n-j}$ cannot hold for all the values of $z_{j-1} > 0, z_j > 0, t > 0$. \square

Thus, we discern that there exist no standard parallel models that can produce i.i.d ICTs. Equivalently, the intersection of the standard serial and standard parallel models is the empty set.

Another similar but perhaps even more intuitive proof of “no intersection of model classes” is the following. Let us observe Lemma 14, if standard parallel models can mimic standard serial models, $P(T_{j+1} + \dots + T_{n'} > t | z_{n'} > z_j)$ is invariant across values of z_j and so must have 0 as the only acceptable value of its derivative with respect to z_j . That is $\frac{S(z_j)S(z_j+t)}{S^2(z_j)} [h(z_j) - h(z_j + t)] = 0$. Now, the only continuous hazard function obeying this condition belongs to the exponential family. However, we already know

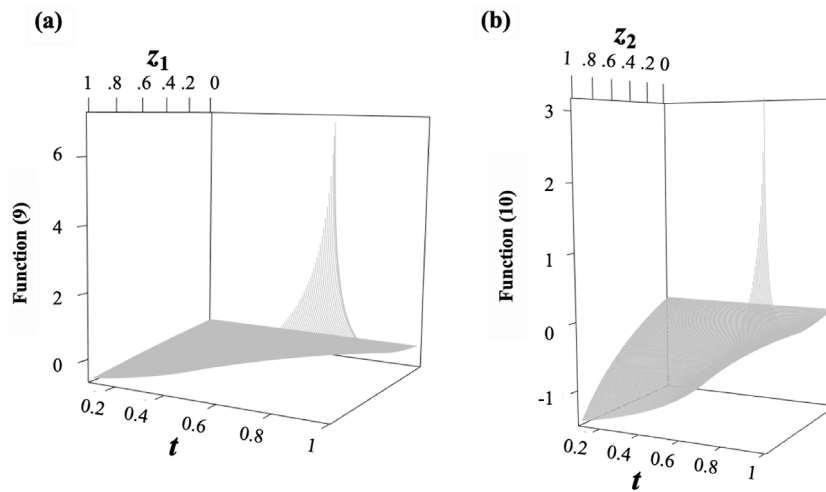


Fig. 9. Plots of (a) function (9) and (b) function (10) for uniformly distributed processors. Note that t , z_1 , and z_2 have the arbitrary and the same unit.

that standard parallel models with exponential distributions possess increasing ICTs (recall McGill's model in the parallel model's representation) rather than the required independently and identically distributed ones of standard serial models. Hence, once again we find that there is only a null intersection of standard serial and standard parallel models.

We thus find an agreeable consequence of this result: Despite the thorny challenge of model mimicry even between the diametrically opposed types of mental architecture, serial vs. parallel, we discover that the prototypical standard serial and standard parallel models do not mimic one another. Therefore, it makes theoretical and empirical sense to employ such critical statistics as ICT and TCT to test between them in psychological environments.

3.4. Implementation of the present theoretical results in the laboratory

It is certainly true that both TCTs as well as ICTs yield important information about underlying information processing systems. Nevertheless, when we move to consider how the present findings might be utilized to investigate psychological phenomena, we discover a curious asymmetry. As already observed, ICTs have been studied, though not voluminously, in the realm of free recall. On the other hand, TCTs have not. Apparently, no one has thought of testing the stochastic independence of TCTs in something like a free recall design. It is certainly possible to do and should, in fact, be done.

Actually, in many experimental milieus where processing issues such as architecture, stopping rule, capacity and so on, have been studied, ICTs and TCTs are fairly invisible. But this could change as researchers begin to explore more complex cognitive realms with tools like SFT.

It is also fact that stochastic independence is often more readily assessed when response frequencies, rather than RTs, are the primary dependent, observable variable. As we continue to unify approaches which have previously been dominated by RTs, such as SFT (e.g., Townsend & Nozawa, 1995) with ones more attached to response frequencies, such as General Recognition Theory (Ashby & Townsend, 1986), our theoretical and statistical power is sure to grow.

Now let us turn to a bit more detailed account of how, at least, our ICT results might relate to the classical free recall paradigm. Therefore, we consider an experimental paradigm in which the subjects recall words from a previously learned list. In the event that the more probable type of standard parallel processing is

present we expect a positive association between the magnitude of ICTs for a word and the position of its recall.

Standard serial models cannot interpret this result as we have discussed earlier. For the recall experiment that we propose currently, standard parallel models can readily account for these data. The experimenter records the RTs of each recalled word. Each RT is counted from the moment the test phase begins to the moment a word is reported. Apparently, each RT is a TCT. Let us label them as $\mathbb{T}_1, \dots, \mathbb{T}_n$. Recall that the TCTs are i.i.d in a standard parallel model. Let us assume each TCT follows the Weibull distribution:

$$f(\mathbb{T}_j) = ku(u\mathbb{T}_j)^{k-1} \exp\left[-(u\mathbb{T}_j)^k\right],$$

where $j \in \{1, \dots, n\}$. The likelihood function, in this case identical to the joint density function, for a standard parallel model can be written as

$$L = f(\mathbb{T}_1)f(\mathbb{T}_2) \cdots f(\mathbb{T}_n).$$

One can use maximum likelihood method to estimate the parameters k and u for the Weibull distribution. We expect that the estimated value of k is not greater than 1, which is consistent with the prediction of Theorem 15 and Corollary 16.

4. Summary and conclusions

With the purpose of rendering this final section independently readable, we drop our previous acronyms.

Our work here differentiates and characterizes the standard multiple-process serial models and the standard multiple-process parallel models by investigating the behavior of (conditional) distributions of the total completion times and survival functions of intercompletion times without assuming any particular forms for the distributions of processing times. We implement this program through mathematical proofs and computational methods. Although the proofs are more complex than with n fixed at $n = 2$, it pleasantly turned out that the major conclusions are in line with the simpler cases.

Thus, we found that for the standard multiple-process serial models and allowing multiple processing orders, positive dependence between the total completion times may or may not hold if no specific distributional functions are imposed on the processing times. In other words, the conditional probability that processors $j + 1, \dots, n$ are completed before some time τ given processors $1, \dots, j$ have already been completed by this time can be greater or not greater than the unconditional (i.e., marginal) probability

that processors $j+1, \dots, n$ are completed by time τ . Interestingly, the prediction for fixed order serial processing reveals that, unlike the situation with mixtures of orders, standard serial models with a fixed processing order are associated with a positive dependence (e.g., covariance > 0) among total completion times.

By contrast, and per definition, for the standard multiple-process parallel models, the total completion times are independent in the sense that the conditional probability that processors $j+1, \dots, n$ are completed before some time τ given processors $1, \dots, j$ have already been completed by this time is equal to the unconditional probability that processors $j+1, \dots, n$ are completed by time τ . According to the different nature of process dependence, one can distinguish a standard multiple-process serial model from a standard multiple-process parallel model.

Moving on to exploration of what happens in standard parallel models with regard to their intercompletion times, we learn that if the hazard function for the processing times is non-increasing, the later stages tend to be successively longer as a function of the magnitude of the earlier intercompletion time, just as when $n = 2$. Moreover, we discover that in standard multiple-process parallel models, the survival functions of the intercompletion times for the later stages (from stage $j+1$ to stage n') conditional on the completion of all the earlier stages is non-decreasing as the processing time for the j th order of processor is increasing if the hazard function of processing times is non-increasing. Recall that j in the term "jth order of processor" represents the position in the processing order.

We further find that the survival function of intercompletion time(s) from stage j to stage $j+1$ is increasing when the ratio of hazard functions is smaller than $\frac{n-j+1}{n-j}$. For all such parallel models the empirical finding in free recall tasks that the intercompletion time grows stochastically with the growth of the number of recalled words is accommodated. Finally, if the cumulative hazard function is concave or linear, the survival function from stage j to stage $j+1$ is increasing.

A limitation of the above theoretical findings is the i.i.d axiom. On the one hand, there is no reason to suspect that as long as a single family of distributions say general gamma, Weibull, Wald, ex-Gaussian and so on, is assumed, that item or position differences should lead to qualitatively different distribution parameters. However, it would be nice to actually demonstrate that fact, as Vorberg and Ulrich (1987) did in the case of the non-i.i.d exponential family. We look forward to a much more rigorous and precise treatment of free recall and related data with the current theory and methodology.

The greatest need appears to be more relevant experimentation. Paradigms associated with free recall as in Bousfield and Sedgewick (e.g., 1944) or Rohrer and Wixted (1994) are an obviously territory ripe for more empirical effort. And, as observed above, there appears to be no obstacle to the analysis of the dependencies of total completion time. Finally, we suspect that there exist other arenas where strategic statistics like intercompletion times and total completion times can prove theoretically beneficial.

Acknowledgments

This study was supported by NSF grant 1331047.

References

- Algom, D., Eidels, A., Hawkins, R. X. D., Jefferson, B., & Townsend, J. T. (2015). Features of response times: Identification of cognitive mechanisms through mathematical modeling. In J. Busemeyer, Z. Wang, A. Eidels, & J. T. Townsend (Eds.), *Oxford handbook of computational and mathematical psychology* (pp. 63–98). New York, NY: Oxford University Press.
- Anderson, J. R. (2013). *The architecture of cognition*. Psychology Press.
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93, 154–179.
- Atkinson, R. C., Holmgren, J., & Juola, J. F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics*, 6(6), 321–326.
- Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology*, 30, 149–165.
- Bousfield, W. A., Sedgewick, C. H. W., & Cohen, B. H. (1954). Certain temporal characteristics of the recall of verbal associates. *American Journal of Psychology*, 67, 111–118.
- Dzhafarov, E. N. (1999). Conditionally selective dependence of random variables on external factors. *Journal of Mathematical Psychology*, 43, 123–157.
- Dzhafarov, E. N. (2003). Selective influence through conditional independence. *Psychometrika*, 68(1), 7–25.
- Dzhafarov, E. N., Schweickert, R., & Sung, K. (2004). Mental architectures with selectively influenced but stochastically interdependent components. *Journal of Mathematical Psychology*, 48, 51–64.
- Egeth, H. E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, 1(4), 245–252.
- Estes, W. K., & Taylor, H. A. (1964). A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences*, 52(2), 446–454.
- Haupt, J. W., Townsend, J. T., & Jefferson, B. (2017). Stochastic foundations of elementary mental architectures. In W. H. Batchelder, H. Colonius, & E. N. Dzhafarov (Eds.), *Modeling and measurement: vol. 2, New handbook of mathematical psychology*. Cambridge, UK: Cambridge University Press.
- Little, D., Altieri, N., Fific, M., & Yang, C. T. (Eds.). (2017). *Systems factorial technology: A theory driven methodology for the identification of perceptual and cognitive mechanisms*. San Diego, CA: Elsevier Academic Press.
- Lohns, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of Memory and Language*, 64(3), 249–255.
- McGill, W. J. (1963). Stochastic latency mechanisms. In R. D. Luce, R. B. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology Vol. 1* (pp. 309–360). New York, NY: Wiley.
- Murdock, B. B. (1971). A parallel-processing model for scanning. *Perception & Psychophysics*, 10(4), 289–291.
- Murdock, B. B., Jr., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Experimental Psychology*, 86, 263–267.
- Newell, A. (1994). *Unified theories of cognition*. Harvard University.
- Patterson, K. E., Meltzer, R. H., & Mandler, G. (1971). Interresponse times in categorized free recall. *Journal of Verbal Learning & Verbal Behavior*, 10, 417–426.
- Pollio, H. R., Kasschau, R. A., & DeNise, H. E. (1968). Associative structure and the temporal characteristics of free recall. *Journal of Experimental Psychology*, 76, 190–197.
- Pollio, H. R., Richards, S., & Lucas, R. (1969). Temporal properties of category recall. *Journal of Verbal Learning & Verbal Behavior*, 8, 529–536.
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1), 129–156.
- Rao, B. L. S. Prakasa (1992). *Identifiability in stochastic models: Characterization of probability distributions*. New York, NY: Academic Press.
- Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory and Cognition*, 22(5), 511–524.
- Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a stroop task. *Journal of Mathematical Psychology*, 18, 105–139.
- Schweickert, R., Fisher, D. L., & Sung, K. (2012). *Discovering cognitive architecture by selectively influencing mental processes*. Singapore: World Scientific Co. Ltd.
- Schweickert, R., Giorgini, M., & Dzhafarov, E. N. (2000). Selective influence and response time cumulative distribution functions in serial-parallel task networks. *Journal of Mathematical Psychology*, 44, 504–535.
- Schweickert, R., & Townsend, J. T. (1989). A trichotomy method: Interactions of factors prolonging sequential and concurrent mental processes in stochastic PERT networks. *Journal of Mathematical Psychology*, 33, 328–347.
- Sederberg, P. B., Gauthier, L. V., Terushkin, V., Miller, J. F., Barnathan, J. A., & Kahana, M. J. (2006). Oscillatory correlates of the primacy effect in episodic memory. *NeuroImage*, 32(3), 1422–1431.
- Sederberg, P. B., Miller, J. F., Howard, W. H., & Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory & Cognition*, 38(6), 689–699.
- Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of human memory*. New York, NY: Academic Press.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1.

- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153(3736), 652–654.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315.
- Townsend, J. T. (1969). Mock parallel and serial models and experimental detection of these. In *Purdue centennial symposium on information processing*. Purdue University: Purdue University Press.
- Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, 10(3), 161–163.
- Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, 25, 168–199.
- Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 133–168). Hillsdale, NJ: Erlbaum Press.
- Townsend, J. T. (1976a). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology*, 14, 219–238.
- Townsend, J. T. (1976b). A stochastic theory of matching processes. *Journal of Mathematical Psychology*, 14, 1–52.
- Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology*, 28, 363–400.
- Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. *Cognitive Theory*, 3, 200–239.
- Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge, UK: Cambridge University Press.
- Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, 39, 321–359.
- Townsend, J. T., & Schweickert, R. (1985). Interactive effects of factors prolonging processes in latent mental networks. In G. d'Ydewalle (Ed.), *XXIII international congress of psychology: vol. 3, Cognition, information processing, and motivation* (pp. 255–276). Amsterdam: North Holland.
- Townsend, J. T., & Schweickert, R. (1989). Toward the trichotomy method: Laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology*, 33, 309–327.
- Townsend, J. T., & Thomas, R. (1994). Stochastic dependencies in parallel and serial models: Effects on systems factorial interactions. *Journal of Mathematical Psychology*, 38, 1–34.
- Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review*, 11, 391–418.
- Townsend, J., Wenger, M. J., & Houpt, J. W. (2018). Uncovering mental architecture and related mechanisms in elementary human perception, cognition and action. In E. J. Wagenmakers (Ed.), *Stevens' handbook of experimental psychology* (pp. 429–454). New York, NY: Wiley.
- Townsend, J. T., Yang, H., & Burns, D. M. (2011). Experimental discrimination of the world's simplest and most antipodal models: The parallel-serial issue. In H. Colonius, & E. Dzhafarov (Eds.), *Descriptive and normative approaches to human behavior in the advanced series on mathematical psychology*. Singapore: World Scientific.
- Vorberg, D., & Ulrich, R. (1987). Random search with unequal search rates: Serial and parallel generalizations of mcgill's model. *Journal of Mathematical Psychology*, 31, 1–23.
- Yang, H., Fific, M., & Townsend, J. T. (2013). Survivor interaction contrast wiggles: Predictions of parallel and serial models for an arbitrary number of processes. *Journal of Mathematical Psychology*, 58, 21–32.
- Zhang, R., & Dzhafarov, E. N. (2015). Noncontextuality with marginal selectivity in reconstructing mental architectures. *Frontiers in Psychology: Cognition*, 1(12), <http://dx.doi.org/10.3389/fpsyg.2015.00735>.
- Zhang, R., Liu, Y., & Townsend, J. T. (2018). A theoretical study of process dependence for the standard two-process serial models and standard two-process parallel models. In T. Lachmann, & T. Weis (Eds.), *Invariances in human information processing* (pp. 117–142). New York, NY: Taylor & Francis Group.