3 Stochastic Foundations of Elementary Mental Architectures

Joseph W. Houpt, James T. Townsend, and Brett Jefferson

3.1 Introduction

In this introductory section, the language will be confined to prose and informality. Later, we will propose a novel general mathematical framework, within which our own as well as a number of other rigorous approaches may be nested.

At the most trite level, serial and parallel systems are different because in the former, sub-processes are taken up one at a time and only exactly when the previous sub-process is completed. In the latter, all sub-processes begin at the same instant, although they may finish at distinct points in time. However, this truly elemental distinction does not prevent mathematical descriptions of the two model classes from often being equivalent (see Townsend, 1972, 1974).

We will first follow the tactics of specific earlier papers as well as Townsend and Ashby (1983), especially Chapters 14 and 15, in delineating fundamental ways in which architectures can be similar and also when they can differ. Subsequently, a unifying and new foundational strategy will be pursued which unifies the homologues and heterologues under a single metatheatrical umbrella. With this strategy in place, we consider possibilities for empirical discriminability of serial and parallel systems or data that allows for model mimicry.

A "system" will be the actual concrete thing itself and of course could be realized in many ways from microchips to neurons. In order to keep the argot

reasonably minimal, we define a "model" to actually be any class of stochastic processes defined by axioms and/or sets of parameters.

Thus, an (admittedly simplistic) parallel system for recognizing pure tones might be a set of tuning forks displayed on a table, each one associated with a specific frequency. A serial rendition of the same kind of task might consist of a sequential exposure of each tuning fork in turn.

Systems possess "processors" which work on the objects fed to them. The "objects" may differ from one another in ways that might affect the time of a processor to complete its individual task. For instance, it has been proposed that the identity of an object could influence the processing time (Townsend, 1976; Snodgrass & Townsend, 1980; Townsend & Evans, 1983; Townsend & Ashby, 1983; Van Zandt & Townsend, 2012), with an important special case being match vs. mismatch of an object to a target item.

The objects to be processed are often associated with one or more dimensions of location. For instance, an auditory set of objects typically possesses a single dimension or order, whereas a visual presentation of letters is usually characterized by a two-dimensional spatial code.

We now turn to a list of informal properties that we associate with the common and distinctive properties of serial and parallel systems. There are two concepts that we use here that will require further formalization later. First, if the representation of an object might be changing in some meaningful way, we say that the object or process is "being worked on." We give a more formal definition of "being worked on" in Definition 3.1. Second, a process is "finished" when there are either no additional changes that will occur in that process or, if additional changes do occur, they will have no effect on the system's response. The formal definition of "finished" is given in Definition 3.3.

1. Properties Held in Common by Parallel and Serial Systems
   - **SP1** The physical properties of the objects can influence processing.
   - **SP2** The order in which previously completed processes finished can influence the processing of unfinished objects.
   - **SP3** The processing times can influence system completion times.
   - **SP4** The order in which the processes finish can influence system completion times.
2. Distinctive Properties of Serial Systems
   - **S1** At any point in time, at most one object is being worked on.
   - **S2** If the processing of an object has not yet started, the properties of that object cannot influence the processing of other objects.
   - **S3** A preordained order of process completions can influence all processes.
3. Distinctive Properties of Parallel Systems
   - **P1** At any point in time, any of the uncompleted objects may be being worked on.
   - **P2** Properties of unfinished objects can influence the processing of other objects through processing interactions.

P3 The order of unfinished process completions cannot influence those processes.

Even before formalizing these properties, it is clear that there are plenty of potential systems that could hold both distinctive properties of serial systems and the distinctive properties of serial systems. For example, **P1** only allows that uncompleted objects may be worked on, but it does not require that they must be worked on, so a system that processes only one unfinished object at time would still satisfy this property. Hence, although it reduces the generality of the systems that may be described within this purview, we also consider a stronger version of **P1**.

P1* At any point in time, any of the uncompleted objects *must* be being worked on.

We will refer to parallel systems that satisfy **P1*** as *strict* parallel systems.

Alternatively, or additionally, we may also consider further constraint on the class of serial systems. For example, the properties outlined thus far allow for a serial system that can start processing an object, but then switch to another process before the first is completed. This would then allow for properties of unfinished objects to influence the processing of other objects, a property that we had nominally reserved for parallel systems. Thus, we consider this further constraint on serial systems,

S4 At any point in time, there is at most one process that has started but is not yet finished.

We will refer to serial systems that satisfy **S4** as *strict* serial systems. Note, along with **S4**, **S2** implies that unfinished objects cannot influence the processing of other objects.

Strict serial models can thus be defined as a distribution on processing order of the objects and a set of distributions for the processing of an object conditioned on that object being the one worked on at a given time (which may be dependent on processing order). Since only one object at a time is being processed, there is no "present status" of processes associated with other objects to depend on, outside of those completed in the past.

A parallel model can be defined as a distribution of processing where the ongoing processing of individual items can potentially depend on the status of any other process or object as well as the history to date of times and order of processing. It is barred from letting the processing order be influenced by subsequent/future order of processing.

We should note that a mental architecture designer, for instance nature, could trade off the ability of a serial system to make its distributions depend on a preordained order, with a dependence on historical events such as previous completion time. In point of fact, both pre-set order and the developing history of completion durations could be influential. Thus, the pre-set order could assign general distributions, such as first is Weibull, second is gamma, and so on. Then, the gamma rate parameter could be determined by the speed of the first stage, with a fast rate

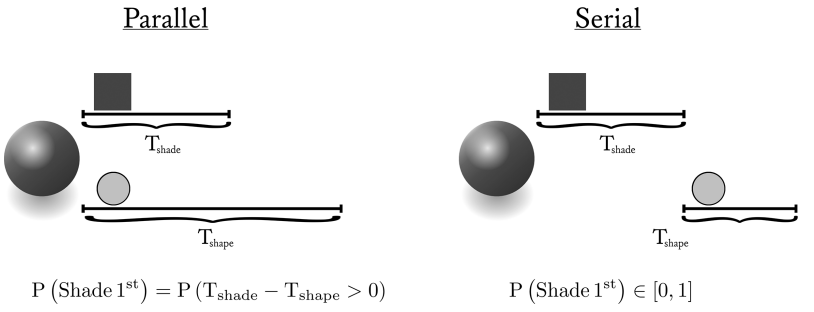

Figure 3.1 *Illustration of a parallel and serial process identifying the shape and shade of an object. For the parallel process, both the shape and shade are worked on at the same time. In the serial process, the shape and shade are worked on at different times. In the parallel process, the probability of the order is a function of the completion time random variables, whereas the only constraint on the probability of the order for the serial process is that it is a probability (i.e., between 0 and 1).*

if the first stage takes longer than 100 ms and a slow rate if it is faster than or equal to 100 ms.

The question of empirical identifiability of the foundational distinction of predetermined order between parallel and serial processes will be discussed in a later section.

The causally acceptable ability of serial systems, but not parallel systems, to preselect the order of processing and processing time distributions (although not the actual realized processing durations) only requires a state space of positions of objects, but not of their composition, similarity to targets and so on. Examples are feature extraction and/or perceptual distance of memory object to a probe item (e.g., see Townsend, 1972). See Figure 3.1, which illustrates this parallel–serial disparity.

On the other hand, the latitude allowed parallel systems to permit various object processes to interact in ways so that states of processing of one (or more) process to depend on that of others calls for a finer-grained description of objects. For instance, feature extraction and computation of similarity to a probe item are examples (e.g., see Townsend, 1972, 1976). See Figure 3.2, which illustrates this parallel–serial disparity.

Our ensuing mathematical account will capture these notions more rigorously. Mathematical modeling departs from statistics and, for the most part, psychometrics, by way of the following precept (following Townsend & Ashby, 1983, chapter 15):

Principle of Correspondent Change

A. Empirical changes in the environment of a stimulating situation should be reflected in a non-vacuous theory or model by corresponding changes or invariances in the model or theory.

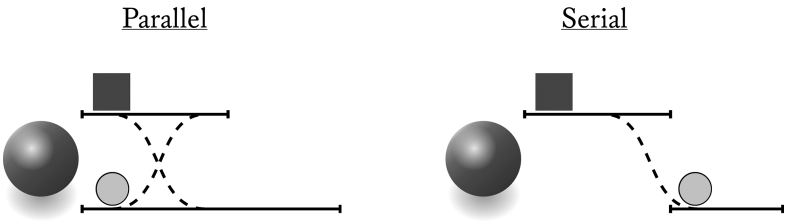

Figure 3.2 *Illustration of interaction in parallel and serial systems. Currently processed information from one source can only affect present or future processing of other sources. Thus, while parallel systems can share information in both directions, serial systems can only share information from earlier processed items to later processed items.*

B. For any given empirical milieu and for any given class of models, there will exist a set of subclasses where models are indistinguishable within their subclass and in that specific milieu.

Principle of Selective Influence

The principle of selective influence (PSI) is a case of principle of correspondent change (PCC) and is likely the most valuable instance to date. The original informal concept was that environmental manipulations can be found which, realized stochastically, slow down or speed up two or more sub-processes (e.g., Sternberg, 1969). Initially, predictions were confined to mean RTs of serial systems, although speculations were made about parallel systems. Additionally, Sternberg (1973) suggested analyses of variance and higher order cumulants. Later, mean RT predictions were offered for more complex systems (e.g., Schweickert, 1978). It also came to be realized that assumptions about how an experimental factor affected an object's response time (RT) distribution were vital to proving claims about parallel vs. serial and more complex architectures, even at the level of mean RTs (Townsend & Schweickert, 1989; Townsend, 1984, 1990; Townsend & Thomas, 1994). Theoretically deep aspects of PSI have been discovered by Dzhafarov and colleagues in recent years (Dzhafarov, 2003; Dzhafarov, Schweickert, & Sung, 2004; Dzhafarov & Gluhovsky, 2006; Kujala & Dzhafarov, 2008; Dzhafarov & Kujala, 2010; Zhang & Dzhafarov, 2015). The concept of selective influence has been generalized by them and others (e.g., Schweickert, Fisher, & Sung, 2012; Algom *et al.*, 2015) to include other dependent variables such as accuracy. The valuable construct of selective influence will be further discussed subsequently.

The PCC is broader still, however, as can be seen in the contrast between the causal confinement of parallel processes to the stochastic processes, dependence on evolving history vs. the ability of a serial system to causally be a function of such future orders. Such a fundamental distinction could, in principle, be associated with observable differences in behavior, whether or not the investigator exploits these differences via manipulation of selective factors.

3.2 General Event Spaces: Formal Probability Theory

The fundamental concept of formal probability theory is that of a probability space. That space is given by a set of elements, usually denoted by Ω, specified subsets of the full space for which a probability is defined, $\mathcal{F}$, and the probability associated with each of those subsets, P. Those triples $(\Omega, \mathcal{F}, P)$ that satisfy the constraints we summarize below are a probability space (also referred to as a probability triple). In light of our discussion above, Ω could be the possible completion times of each of the processes, $(\mathbb{R}^+)^n$ for n processes. Alternatively, Ω could be the possible states of each of the processes at any time, $\mathbb{R}^n \times \mathbb{R}^+$, where each of the dimensions of $\mathbb{R}^n$ corresponds to the activation state of a process and $\mathbb{R}^+$ corresponds to time since stimulus onset.

The subsets in $\mathcal{F}$ can include any subsets of Ω that would normally be useful for distinguishing serial and parallel processing, but not all subsets of Ω are required to be in $\mathcal{F}$. $\mathcal{F}$ must include the entire space Ω and its complement, the empty set $\emptyset$. Also, for any set $A \in \mathcal{F}$, the complement A^c must be in $\mathcal{F}$. $\mathcal{F}$ must also be closed under countable unions and countable intersections. Thus, if $A_1, A_2, \ldots$ are in $\mathcal{F}$, then $\cup_{i=1,2,\ldots} A_i$ and $\cap_{i=1,2,\ldots} A_i$ must also be in $\mathcal{F}$. When a collection of sets satisfies these constraints it is called a σ-algebra or σ-field.

For modeling cognitive processing times, we want to be certain that intervals are included in $\mathcal{F}$. For example, we will need to refer to probabilities such as "the probability that process A takes less than one second and process B takes between two and three seconds." The smallest σ-algebra on the real numbers that includes all intervals is called the Borel σ-algebra. That this σ-algebra exists follows from the famous Kolmogorov extension theorem (e.g., Billingsley, 1995, §3).

For P to be a probability measure, the probability of the empty set must be zero, $P(\emptyset) = 0$, the probability of the whole space must be one, $P(\Omega) = 1$, and P must be countably additive. Countable additivity means that for any *disjoint* sets, $A_1, A_2, \ldots$ in $\mathcal{F}$, the probability of the union of the sets (which must be in $\mathcal{F}$) is the sum of the probability of each set, $P(\cup_{i=1,2,\ldots} A_i) = \sum_{i=1,2,\ldots} P(A_i)$.

In general, any probability measure that satisfies these conditions could be used for modeling mental processing. In practice, models tend to take one of two forms. Many modelers will specify the probability measure by associating the random events with commonly used distributions, such as an exponential distribution or a more complex distribution based on processing assumptions, such as the linear ballistic accumulator (Brown & Heathcote, 2008) or the drift-diffusion model (Ratcliff, 1978). Alternatively, one can study systems by determining what inferences can be made about the systems using only minimal additional constraints on the probability measure (c.f. Townsend & Nozawa, 1995; Dzhafarov *et al.*, 2004; Zhang & Dzhafarov, 2015).

We mentioned random variables and distributions, with which the reader will no doubt be at least minimally familiar, but these terms also have a more rigorous definition within probability theory. A univariate random variable X is a measurable function from the space Ω to the real numbers. To be measurable, the preimage of

every Borel set under X must be in the σ-algebra, $\mathcal{F}$, on Ω, i.e., for any Borel set B, $X^{-1}(B) \in \mathcal{F}$.

Suppose, for example, that instead of working with a space Ω that is defined on the response times, which are elements of $\mathbb{R}$, Ω is the space of a participant's actions in an experiment. Then the elements of Ω are not real numbers, but actions. To discuss response times, we would need to map from the space of actions to the real number representing the time it takes for an action. The measurability requirement allows one to work with probabilities of either the response time or the actions because there are connections between the measurable sets $\mathcal{F}$ in Ω and measurable (specifically Borel) sets in $\mathbb{R}$.

The distribution, μ, of a random variable X is the function on the Borel sets that gives the probability measure of the corresponding set in $\mathcal{F}$ given by X. Thus, if B is a Borel set, then $\mu(B) = P(X^{-1}(B)) = P(X \in B)$ where P is the probability measure from the space on which X is defined.

A common practice in RT modeling is to assume that response times have a familiar distribution, such as a gamma or Weibull distribution. This approach sets the measure on the Borel sets on the real line rather than setting the measure on a separate probability space of interest. In many cases, researchers are only interested in direct statements about the response time distributions, so there is no loss in modeling them directly with a distribution. As we see below, there is some danger in ignoring the complexity of the processes that lead to the response time when considering the effect of experimental variables on the response time distributions.

A single random variable can be quickly generalized to a multivariate random variable or random vector. If X is an n-dimensional random vector, then each dimension is a random variable on the same probability space. That is, each X_i is a measurable function from the probability space to the real numbers. The random vector has a probability distribution given by $\mu(B) = P(X \in B) = P(X_1 \in B_1, \ldots, X_n \in B_n)$ when B is an n-dimensional Borel set. This measure μ is the joint distribution of the random variables $X_1, \ldots, X_n$. Because each X_i is a random variable, we also have the marginal distributions, $\mu_i(B_i) = P(X_i \in B_i)$.

From multivariate random variables, we can generalize to stochastic processes. These are essentially multivariate random variables, but with a possibly infinite, even continuous, index set. Formally, a stochastic process can be defined by the set of random variables $\{X_t\}_{t \in T}$ on a probability space $(\Omega, \mathcal{F}, P)$, where T is a possibly infinite index set, such that for any finite subset $A \subset T$, $\{X_t\}_{t \in A}$ is a multivariate random variable.[1]

Most commonly, a stochastic process is a process that unfolds over time so we may be interested in the future values, $\{X(t), t > s\}$, given the value at a specific time, $X(s)$. In this case, we know more about the random variables as time progresses. For example, if we were using a stochastic process to model the movement

1 To be able to add constraints to stochastic processes that depend on an uncountable subset $S \subset \{X(t)\}_T$, we need a more complex definition. While these constraints are important (e.g., continuity of the sample paths), we avoid the additional detail here for the sake of brevity.

of a mouse cursor across the screen, then at a given time t we would know where the mouse had been at every time up until t, i.e., if $X(s)$ is the cursor position at time s, then we could have observed every $X(s)$ for $s \leq t$. The formal way of describing this increase in knowledge over time is by specifying a collection of σ-fields that increase over time: for all t there is a $\mathcal{F}_t \in \mathcal{F}$ such that for any $s \leq t, \mathcal{F}_s \subset \mathcal{F}_t$. This type of collection of σ-fields is known as a filtration. We can capture an increase in knowledge about a stochastic process over time by requiring that the random variables $X(t)$ are each $\mathcal{F}_t$ measurable where the collection of $\mathcal{F}_t$ is a filtration.

Stating that $X(t)$ is measurable $\mathcal{F}_t$ means that the possible events of the system $\mathbf{X}$ at time t are included in the σ-field $\mathcal{F}_t$. Given a system $\mathbf{X}$, there is a naturally induced filtration constructed as the union of the smallest σ-fields for each t,

$$\mathcal{F}_t = \bigcup \left\{ X(s)^{-1}(B) : B \text{ is a Borel set} \right\}_{s \in [0,t]}.$$

For the induced filtration, it is clear that given $\mathcal{F}_t$ we have complete knowledge about $X(t)$, however, we do not necessarily know anything about $X(t')$ for $t' > t$.

Example 3.1 Consider a single process $X(t)$ defined at discrete points in time $t \in \mathbb{N}$. For each t, let $X(t)$ be continuous random variables defined on $\mathbb{R}$. Then each $\mathcal{F}_i$, where $i \in \mathbb{N}$, can be thought of as the disjoint union of i copies of Borel sets. So one event of $\mathcal{F}_2$, for example, is the disjoint union $(2, 26) \sqcup (1, 5.56)$. An event of $\mathcal{F}_{40}$ could be the disjoint union of 40 copies of $(0, 1)$!

Note two things from this example. First, the notion of a filtration is separate from the individual probabilities associated with each $X(t)$, although the example makes no mention of any specific distribution. Second, we could define $X(41)$ to be a discrete random variable. Doing so would make it clear that $\mathcal{F}_{40}$ does not imply anything about the subsets/events of $\{X(41)^{-1}\}$. Generally, it is useful to think about each $\{X_i(t)\}_t$ as independent continuous random variables.

The next important concept is of a conditional probability. Suppose we have random variables X and Y on the measure space $(\Omega, \mathcal{F}, P)$. In many cases we want to examine the properties of X given knowledge of Y (for example, the response times from trials when a person is correct). As long as the event we are conditioning on has positive probability, then the traditional definition is sufficient, $P(X|Y) = P(X, Y)/P(Y)$. However, if $P(Y)$ is zero, then this would suggest that $P(X|Y)$ is undefined or infinite. That might work if you were conditioning on correct or incorrect when there is no chance of the participant being incorrect because we would never worry about conditioning on something that is impossible. But what about conditioning on a particular response time? If a response time could be any positive real number (or even any positive real number less than some upper bound) then the probability that it is a specific value would have to be zero for most values. Nonetheless, the response time must be some value. Suppose that we observe a response time of 326 ms. In the conditional $P(\text{Correct} \mid \text{RT} = 326\,\text{ms})$ the probability of $\text{RT} = 326\,\text{ms}$ is zero.

To deal with the issue of conditioning on measure zero events, conditional probabilities are defined as a function of the conditioned variable, $f(B) = P(Y|X \in B)$. For this function to make sense, it needs to satisfy two requirements. First, the function needs to be measurable with respect to X (actually $\sigma(X)$: the σ-algebra generated by X). Second, we want to preserve the equality $P(X, Y) = P(X)P(Y|X)$. Formally, we require that for all $B \in \sigma(X)$, $\int_B P(Y|X \in B)\, dP = P(Y, X \in B)$ where the dP refers to the measure P on the probability space.

The next concept from the theory of stochastic processes that we will need is that of a stopping time. A stopping time is a random $\tau \in T$ such that $\{\tau \leq t\} \in \mathcal{F}_t$ for all t in the index set T. For example the stopping time could be a fixed time, e.g., $\tau = 326$ ms, or it could be based on some event that is measurable with respect to $\mathcal{F}$, e.g., τ is the first time that $X_i(t) \geq 10$. For those familiar with information accumulator models, they often assume information is accumulated up to some threshold, then a response is made. In this case the response time is a stopping time for the σ-field generated by the amount of information accumulated at each time t. When discussing the exhaustive and first-terminating stopping rules for mental processes in the next section, we will make use of the fact that if τ and ν are both stopping times for a given σ-field, then so are the minimum and the maximum of τ and ν.

In addition to the stopping times for processes, we will also be interested in the order in which processes complete. For example, if the processes X_1, X_2 and X_3 complete at T_1, T_2 and T_3, respectively, we may be only interested in whether $T_2 < T_1 < T_3$ or $T_1 < T_2 < T_3$, etc. To refer to the first process to complete, when we are not concerned with which process it is, we will use $X_{(1)}$ and similarly $T_{(1)}$ for the completion time of the first process to finish and $t_{(1)}$ for a particular observed value of the completion time of the first process to be finished.

With these fundamental concepts in mind, we now turn to a formalization of the elementary properties of mental processes outlined in the introduction, using stochastic processes.

3.3 Establishing Serial and Parallel Distinctions through the σ-Spaces

We begin by formalizing the properties of serial and parallel systems that we outlined above. To do so, we connect the informal idea of "processes" developed above with the mathematical concept of adapted stochastic processes. If there are n elements of information that the cognitive system is working with, then we need a stochastic process corresponding to each of those elements, $\{X_i(t), t \in [0, \infty)\}_{i=1\ldots n}$. Equivalently, we can represent these processes as an n-dimensional stochastic process $\mathbf{X}(t), t \in [0, \infty)$. In this case, we can call each X_i a sub-process. We also want this process to be adapted to preserve causality. Our main goal is that future events do not change the present (i.e., $\mathbf{X}(t)$ is determined by events in the past $s < t$ or in the present, but not the future $s > t$). For this goal,

it is sufficient that t is ordered in the sense that $\mathbf{X}(t)$ is measurable with respect to the filtration $\mathcal{F}_t$. To formalize the property that a serial system may have a predetermined order of processing, we allow that the order is measurable $\mathcal{F}_0$, i.e., that the order might be set at $t = 0$ or some $t < 0$.

Example 3.2 For concreteness, suppose a person is asked to judge the size, weight, and color of a ball. A reasonable assumption is that for each dimension of the stimulus, the associated sub-process has the form of some information accrual process such as $X_i(t) \overset{d}{\sim} unif(t_{i_1}, t_{i_2})$ for each t_{i_1} and t_{i_2} functions of t. The simplest case is to allow the $t_{i_1} = 0$ and $t_{i_2} = t$. Here, as time passes, the probability of acquiring more information about each dimension grows linearly. Naturally, we could impose some dependence across sub-processes, but for illustrative purposes consider the independent case. We will refer to this example as needed throughout the chapter.

The first distinguishing property of serial and parallel systems concerns "being worked on." We can formalize "being worked on" as follows:

Definition 3.1 Process i is being worked on at time t if there is no $s > t$ such that for all $v \in (t, s]$,

$$P\left(X_i(t) = X_i(v)\right) = 1.$$

Less formally, for any amount of time that passes, there is a chance that X_i changes state during that time. Similarly, if the process is not being worked on, there is some amount of change in time for which there is no chance that the state will change. If color processing is delayed as above, then color is not being worked on for all $t \in [0, T_1)$. Note that, according to this definition, if X_i changes state at deterministic intervals, then X_i is not being worked on during the interval. This may seem odd for a discrete time system, but this can be mitigated by assuming t, s, and v are positive integers in the definition.

Next, we need definitions for starting and completing processing.

Definition 3.2 A sub-process $X_i(t)$ has started if for any $s \leq t$, X_i was being worked on at s. When the minimum (or infimum) of such s exists, s', we say that X_i started at s'.

In Example 3.2, we could delay processing of color until size has completed. This would be realized as

$$X_3(t) \overset{d}{\sim} \begin{cases} 0 & \text{if } t \leq T_1 \\ unif(0, t) & \text{if } t > T_1. \end{cases}$$

The processing of color hasn't started until time T_1.

To formalize process completion, we use the first passage (or first exit) time concept.

Definition 3.3 Let each sub-process $X_i(t)$ have a predetermined completion region, $\theta_i \subset [0, \infty)$. We will say that $X_i(t)$ is complete (or finished) if $t \geq \tau =$

$\min_s X_i(s) \in \theta_i$. For the system completion time, we have the completion region $\Theta \subset \mathbb{R}^n_+$ and say that the system has completed if $t \geq \min_t \mathbf{X}(t) \in \Theta$. A system may need all sub-processes to be complete (exhaustive system), or it may need only a subset of the processes to be complete (referred to as a self-terminating system).

Example 3.3 From Definition 3.3, we can consider the three sub-process system $\mathbf{X} = \{X_1, X_2, X_3\}$ described above with completion regions $[1, 2]$, $[3, 4]$, and $[5, 6]$, respectively. An exhaustive system can be constructed by defining $\Theta = [1, 2] \times [3, 4] \times [5, 6]$. All sub-processes must be in their respective completion regions simultaneously in order for processing to complete. Likewise, a self-terminating system can be constructed by defining

$$\begin{aligned}\Theta = &[1, 2] \times (-\infty, \infty) \times (-\infty, \infty) \\ &\cup (-\infty, \infty) \times [3, 4] \times (-\infty, \infty) \\ &\cup (-\infty, \infty) \times (-\infty, \infty) \times [5, 6],\end{aligned}$$

the union of infinite strips. Here any sub-process may enter its respective completion region for processing to complete.

One may wonder what happens after completion. Perhaps the simplest assumption would be that processing stops, i.e., if τ_i is the completion time of X_i, then for all $t > \tau_i$, $X_i(t) = X_i(\tau_i)$. This has the advantage of connecting channel completion times to system completion times with simple rules (e.g., the system completion time of an exhaustive system is the maximum of the sub-process completion times). In some cases, requiring processing to stop upon completion leads to more constraints, which may be unintuitive. For example, it precludes the possibility that a participant might second-guess himself after responding. At the other end of the spectrum, we could leave the process totally unconstrained, free to move in and out of the completion region and free to have any continued processing influence unfinished processes. For convenience, we take the middle ground and assume that, although processing may continue after completion, the outcome of the (sub-) process does not change and changes in sub-processes after they have completed cannot influence other processes.

We now augment the framework to codify the properties that we outlined that both serial and parallel systems should satisfy. We follow Dzhafarov (2003) by limiting the probability space associated with a perceptual process to the internal states of the system and assume that the environment in which the system operates is deterministic. This assumption is not necessarily required for our theory, but it will simplify the connection between these systems and the theory of selective influence later in the chapter. Hence, we identify the probability space with a particular configuration of the environment, i.e., if ζ indicates the specifics of the environment, then the relevant probability space is $\left(\Omega_\zeta, \mathcal{F}_\zeta, P_\zeta\right)$. This allows for each of the processes defined on the space to be different under different environmental conditions, and in particular, satisfies **SP1**. In fact, this formalization

requires that the processes are different under different configurations of the environment because equality (and almost-sure equality) in measure theory are defined with respect to the underlying measure space. It may be possible that there are two different probability spaces that nonetheless lead to exactly the same distributions for $\mathbf{X}_t$ for all t even when the two event spaces are not equivalent. Hence, in keeping with **SP1**, we also require that the environmental conditions are allowed to lead to different distributions on $\mathbf{X}_t$.

Property SP1 Let ζ_a and ζ_b be any two possible states of the world. Then it is possible that $\left(\Omega, \mathcal{F}_{\zeta_a}, P_{\zeta_a}\right) \neq \left(\Omega, \mathcal{F}_{\zeta_b}, P_{\zeta_b}\right)$.

Property SP2 Let $O(I)$ be the order of T_i for $i \in I$. If $I_t = \{i : T_i < t\}$ then $\sigma\left(O(I_t)\right) \subset \mathcal{F}(t)$.

Property SP3 For all $i, \sigma\left(T_i\right) \in \mathcal{F}_T$.

Property SP4 Let $O(I)$ be the order of T_i for $i \in I$. If I is the set of all sub-processes, then $\sigma\left(O(I)\right) \subset \mathcal{F}_T$.

S1 states that a serial system can only have one process being worked on at any given time. Formally, this can be stated as follows.

Property S1 In a serial system, if, for any i, there is no $s > t$ such that for all $v \in (t, s], P\left(X_i(t) = X_i(v)\right) = 1$ then for all $j \neq i$, there exists some $s > t$ such that for all $v \in (t, s], P\left(X_j(t) = X_j(v)\right) = 1$ (or, equivalently, $P\left(X_j(t) \neq X_j(v)\right) = 0$).

Property S2 Let ζ_a and ζ_b be any two possible states of the world such that for any process i that has been worked on by t, object i is the same in both ζ_a and ζ_b. Then, in a serial system, $\mathcal{F}_{t,\zeta_a} = \mathcal{F}_{t,\zeta_b}$ and $X_a(t)$ defined on $\left(\Omega, \mathcal{F}_t, \Pr_{\zeta_a}\right)$ is equal in distribution to $X_b(t)$ defined on $\left(\Omega, \mathcal{F}_t, \Pr_{\zeta_b}\right)$.

Property S3 Let $O(I)$ be the order of T_i for $i \in I$. In a serial system, if I is the set of all sub-processes and for all $t > 0$, it is possible that $\sigma(O(I)) \subset \mathcal{F}(t)$.

Consider again Example 3.1. Now suppose that the order of serial processing was always color first. Then $\mathcal{F}_0$ would contain two events,

$\mathcal{F}_0 = \{\{\text{color first, size second, weight third}\}, \{\text{color first, size third, weight second}\}\}$.

Property S4 In serial systems, for all t, if there exists i such that $0 < X_i(t) < \theta_i$, then for all $j \neq i$, either $X_j(t) = 0$ or $T_j < t$.

Property P1 In parallel systems, for all t, and all i such that $T_i > t$, it is possible that there is no $s_i > t$ such that for all $v \in (t, s_i], P\left(X_i(t) = X_i(v)\right) = 1$.

Property P1* In strict parallel systems, for all t, and all i such that $T_i > t$, there is no $s_i > t$ such that for all $v \in (t, s_i], P\left(X_i(t) = X_i(v)\right) = 1$.

Property P2 In parallel systems, for any object i under the state of the world ζ_a such that $T_i > t$, if there exists another object j such that $T_j > t$, and another state of the world ζ_b in which there is an object j_b identical to object j from ζ_a, then it is

possible that $\mathcal{F}_{t,\zeta_a} \neq \mathcal{F}_{t,\zeta_b}$ and in particular that $X_j(t)$ defined on $\left(\Omega, \mathcal{F}_t, \Pr_{\zeta_a}\right)$ may not be equal in distribution to $X_{j_b}(t)$ defined on $\left(\Omega, \mathcal{F}_t, \Pr_{\zeta_b}\right)$.

Property P3 Let $O(I)$ be the order of T_i for $i \in I$. In parallel systems, if $I = \{i : T_i > t\}$, then $\sigma\left(O(I)\right) \nsubseteq \mathcal{F}(t)$.

When **S4** is assumed, i.e., strict serial systems, then **S1** implies that the identities of unfinished objects that are not being worked on cannot influence the processing of other objects. Here, we use the filtration concept to encode the idea of what has influence on an outcome. If object i is unfinished and not being worked on at time t, then the processing of all objects $j \neq i$ at time t cannot be influenced by object i.

Thus any set in $\mathcal{F}_t$ that one could condition upon when measuring the probabilities associated with X_j contains no information about X_i. In parallel systems, it is possible for information about unfinished processes to influence other processes, so there is no such constraint on the filtrations.

3.4 Causality in Parallel and Serial Systems

Although the current definitions are similar in many ways to the definitions presented in Townsend and Ashby (1983), the use of filtrations and adaptations are new. Townsend and Ashby (1983) defined models of parallel and serial systems as probability triples associated with the sample space of sub-processes' completion times and orders, so they did not investigate the notion of differences in the filtrations associated with parallel and serial models. We are engaging the notion of filtrations to achieve our goals of more formally constraining the definitions of parallel and serial systems with notions of causality.

In particular, for the models to maintain standard notions of causality, future events should have no influence on past events. For example, the total system completion time (i.e., exhaustive processing of all objects) should not influence the completion time of the first of many sub-processes to complete. In terms of adapted stochastic processes, the overall system completion time is not in the adapted σ-algebra at the first sub-process's completion time. For example, if there are three sub-processes but only the first has finished ($T_1 < t < T_2$), then the system completion time is not measurable $\mathcal{F}_t$, i.e., it does not make sense to talk about probabilities such as $P(T \in [a, b])$ or $P(T \in [a, n]|T_1 = \tau)$. However, if $t \geq T$, then both the system completion time and the first sub-process completion time are in the filtration $\mathcal{F}_t$, which allows us to consider probabilities such as $P(T \in [a, b]|T_1 = \tau)$ or $P(T_1|T = s)$.

3.5 Experimental Identifiability of Architectural Distinctions of Causal Mental Systems

We now turn our attention to focus on the ways in which the underlying event spaces (i.e., the σ-fields) contrast between serial and parallel models. These distinctions differ in their empirical consequences. Some lead to predictable

and potentially testable differences even within very elementary experimental conditions, whereas others require more subtle treatments.

The distinguishing characteristics given in the introduction and formalized in Section 3.3 are not alone sufficient for the systems to always be distinct, because each of the constraints on one system (e.g., that serial systems can have at most one process being worked on) is only stated as something the other is allowed, but not required, to do (e.g., in a parallel system, any unfinished process being worked on). This leads to sufficient conditions for rejecting one class of system or the other, but in cases where the constraint is satisfied, we are not able to distinguish between the models. For this reason, we have also introduced a strict version of **P1** to allow for distinguishing between the two classes of systems.

First, suppose we can observe the filtration $\mathcal{F}_t$ for all t at which a process completes. As long as at least two sub-processes have not yet finished, then a serial system is allowed to have more information about the order of processing in each of those $\mathcal{F}_t$ than a parallel system. For example, if the serial system determines the complete order immediately once processing begins, then the order is in $\mathcal{F}_t$ for all $t > 0$. Because a parallel system cannot predetermine the order of processing, then whenever there are still two unfinished processes, their order is not yet in $\mathcal{F}_t$. Once all but one of the processes has completed, that unfinished process will be the last, so the order of processing is determined (i.e., in $\mathcal{F}_t$). Because there is no requirement that serial systems *must* determine the full order before processing begins, this is merely a sufficient condition for rejecting parallel systems.

Suppose instead that we have access to the complete distribution over the processes when the system is finished, i.e., the distribution of $\mathbf{X}(s)$ for all s up to some t greater than or equal to the total completion time of the system. In terms of filtrations, we have full knowledge of the distributions of random variables $\mathbf{X}$ that are measurable $\mathcal{F}_t$ for some t greater or equal to the system completion time. We want to emphasize that this information is about the probabilities associated with different states of the system, not about particular sample paths that the system actually follows. At this point the order of completion is measurable, whether the system is parallel or serial, and hence the presence of order in the σ-field is no longer a distinguishing factor. Nonetheless, based on this strong assumption, we could easily distinguish between strict parallel and serial systems by conditioning on the event that a particular object is being worked on at a given time and check if any other processes are being worked on.

If, instead of accessing properties of the filtration and/or distributions, we only have access to sample paths of the system, we can check sample path properties to potentially reject serial systems.

Proposition 3.1 *The derivatives of the sample path of at most one sub-process from a serial system can be either be non-zero or non-existent at any $t > 0$.*

Proof Consider a sample path x_i of X_i that is not being worked on at time t. By the definition being worked on, there exists some $s_i > t$ such that for all $v \in (t, s]i$, $x_i(v) - x_i(t) = 0$ almost surely. Hence, for any $v \in (t, s)$, the derivative

of $x_i(v)$ exists and is zero almost surely. Now consider $s = \min_i s_i$ for all x_i that are not being worked on at t (which exists as we are assuming a finite number of sub-processes). Then the derivatives of the sample path of each of those sub-processes exists and is zero for some interval after t. Because, in a serial system, there is at most one sub-process being worked on at any t, there is an interval after any t for which at most one sub-process's sample path is either non-zero or non-existent. □

In contrast, in a strict parallel system, regardless of the interaction among processes, there is some possibility that all of the unfinished processes change within an arbitrarily small time interval of each other.

Proposition 3.2 *In a strict parallel system, there is positive probability that for some $t > 0$ and any $\epsilon > 0$,*

$$X_1(t+\epsilon) - X_1(t) > 0 \text{ and } X_2(t+\epsilon) - X_2(t) > 0.$$

Proof Strict parallel systems must have all processes either being worked on or finished at all times, so for any time $t > 0$ and $\epsilon > 0$, if X_i are the sub-processes that have not yet finished, the probability $P[X_i(t+\epsilon) - X_i(t) = 0] \neq 1$. Hence, for all i, $P\left[X_i(t+\epsilon) - X_i(t) \neq 0\right] > 0$. □

In practice, we may observe a sample trajectory on which no sub-processes change within a given ϵ of one another. Nonetheless, if we continue to observe more sample trajectories, we would eventually observe one or more processes changing within the same ϵ interval, no matter how small the ϵ.

Next, we show that parallel and serial systems cannot be distinguished based on the system completion times under a single experimental condition. Recall that the system completion time distribution is only meaningful for $\mathcal{F}_t$ in which t is larger than system completion time.

Proposition 3.3 *Suppose we only have access to the distribution over completion time (and hence also order) of the sub-processes in the system.*[2] *Without further restrictions, serial system and strict parallel systems cannot be distinguished at this level.*

Proof Consider an arbitrary serial process with distributions $P(\mathcal{O})$ on the order of processing and $f(t_1, t_2, \ldots, t_n|\mathcal{O})$ on the time it takes to complete each sub-process. Following Townsend and Ashby's (1983) notation, we let the random vector $(A_1, A_2, \ldots, A_n)$ be the vector in which each element gives the ordinal rank in which that item finished. Then a parallel process can completely mimic the serial distribution, f, by setting the completion time distribution of each object in the parallel system equal to the sum of the completion time of each object that completed before it along with the objects' completion time in the serial system. That is, for every possible outcome $\omega \in \Omega$,

2 This is the same case that Townsend and Ashby (1983) examine in Chapter 14.

$$T_i^p(\omega) = T_i^s + \sum_{j:A_j(\omega)<A_i(\omega)} T_j^s.$$

In particular, this means that the probability over orderings are the same for both models.

We can use a similar mapping in reverse to arrive at a parallel system starting from a serial system,

$$T_i^s(\omega) = T_i^p(\omega) - T_{j:a_i-a_j=1}(\omega).$$

Under this reverse mapping, the density over orderings is given by

$$p = \int_0^\infty \int_{t_{(1)}}^\infty \cdots \int_{t_{(n)}}^\infty g\left(t_{(1)}, t_{(2)}, \ldots, t_{(n)}\right)\, d\tau_{(n)} \ldots d\tau_{(1)}.$$

This mapping is quite general, so if the only observables are the intercompletion times T (and hence the orderings), serial and parallel systems can be nearly non-discriminable. Consider the system at the completion of the first process, $t_{(1)}$. Under a serial system, the entire order of completion may be in the measure space at $t_{(1)}$ and the ordering of later processes can influence $t_{(1)}$. This is, of course, impossible in a parallel system. Unfortunately, if we cannot observe the ordering at $t_{(1)}$, then there is no way to test whether future completion orderings have an effect on $t_{(1)}$. Once we have observed the completion order, we could evaluate the conditional probabilities,

$$P(\tau_1 \in \mathcal{T} \mid \langle A, B, C\rangle) \stackrel{?}{=} P(\tau_1 \in \mathcal{T} \mid \langle A, C, B\rangle).$$

However, by the above mapping, whichever pattern of dependence we observe in a serial system can be mimicked by a parallel system. □

Although the mapping holds in general, it can lead to some unintuitive consequences. First, for a serial system, the probability of process a finishing first is the sum of the density when a is first under order $O_1 = a, b, c$ plus the density when a is first under order $O_2 = a, c, b$. If we assume within-stage independence in the parallel system (a system is within-stage independent if, for any interval of time during which no sub-process completes, all sub-processes that have not yet finished are independent on that interval), then for a parallel system to perfectly mimic this system,

$$\begin{aligned}
&P(\mathcal{O}_1) f_{a_1}(t_{a_1}|\mathcal{O}_1) + P(\mathcal{O}_2) f_{a_1}(t_{a_1}|\mathcal{O}_2) \\
&= P\left(T_{a_1} = t_{a_1} < T_{b_1}, T_{c_1}\right) P\left(T_{b_2} < T_{c_2}\right) + P\left(T_{a_1} = t_{a_1} < T_{b_1}, T_{c_1}\right) P\left(T_{c_2} < T_{b_2}\right) \\
&= g_{a_1}(t_{a_1}) G_{b_1}(t_{a_1}) G_{c_1}(t_{a_1}) \left(\int_0^\infty g_{b_2}(t_2) G_{c_2}(t_2)\, dt_2 + \int_0^\infty g_{c_2}(t_2) G_{b_2}(t_2)\, dt_2 \right).
\end{aligned}$$

This implies that distribution of the completion time of a when it is first in the mimicking parallel system is given by

$$G_{a_1}(t) = \exp\left\{ -\int_{t_{a_1}=0}^{t} \frac{P(\mathcal{O}_1) f_{a_1}(t_1|\mathcal{O}_1) + P(\mathcal{O}_2) f_{a_1}(t_1|\mathcal{O}_2)}{P(\mathcal{O}_1) F_{a_1}(t_1|\mathcal{O}_1) + P(\mathcal{O}_2) F_{a_1}(t_1|\mathcal{O}_2) + \cdots + P(\mathcal{O}_6) f_{c_1}(t_1|\mathcal{O}_6)} \, dt_{a_1} \right\}.$$

Hence, while it may seem reasonable to assume that the serial system where if a is first, its completion time has a gamma distribution, if b is first it has a Weibull distribution and if c is first it has a truncated normal distribution, the mimicking parallel system would be required to have a mixture of those distributions for the completion time of a when it finishes first. Nonetheless, there is nothing theoretically preventing a parallel system from having such a distribution.

3.5.1 Distinguishing Parallel and Serial Systems with Selective Influence Manipulations

One limitation of this early methodology was that the mathematical underpinnings that justified the proposed experimental inferences were lacking. Although it had long been known that the expectation of additive random variables, say $T_1 + T_2$, would be an additive function, Sternberg was aware that α_1 might non-selectively affect T_2, or the other way around. However, it was later shown that even with α_i affecting the proper sub-process directly if, say, T_2 was correlated with T_1 (e.g., T_2 tends to be faster if T_1 was slow and vice versa), then selective influence at the level of the mean RT would usually fail (e.g., Townsend, 1984).

More importantly, if selective influence was assumed to take place at a stronger level, for instance, ordering the distributions such that, say, $F_1(t_1|\alpha_1) > F_1(t_1|\alpha_1^*)$ if and only if $\alpha_1 > \alpha_1^*$, then other architectures could also be tested, in particular parallel architectures (Townsend, 1984, 1990). Furthermore, diverse architectures could be assessed at the far more powerful level of distributional functions (Townsend & Nozawa, 1995; Schweickert, Giorgini, & Dzhafarov, 2000; Dzhafarov *et al.*, 2004).

An important issue has long been how and when selective influence, in the sense of invariance of the marginal distribution of a random completion time (e.g., T_2) marginalized over the other random times (e.g., T_1) might be realized (e.g., Townsend & Ashby, 1983, chapter 11) if sheer stochastic independence was not in force. Although a completely global answer to this question has not yet appeared, Dzhafarov (2003) proposed that the definition of selective influence be as follows:

Definition 3.4 The factors $\alpha_1, \alpha_2, \ldots, \alpha_n$ selectively influence the random variables $X_1, X_2, \ldots, X_n$ if there exists a set of independent random variables $C, S_1, S_2, \ldots, S_n$ that do not change as a function of any α_i, and measurable functions $\lambda_1, \lambda_2, \ldots \lambda_n$ such that

$$X_1 = \lambda_1(\alpha_1, S_1, C), X_2 = \lambda_2(\alpha_2, S_2, C), \ldots X_1 = \lambda_n(\alpha_n, S_n, C).$$

The reader may observe that although the X_1 and X_2 in the definition are not independent, they are conditionally independent (given C). A key consequence

of this conditional independence is that the major theorems on testing of various architectures (Townsend & Nozawa, 1995; Schweickert *et al.*, 2000; Houpt & Townsend, 2011) go through unimpeded (Dzhafarov *et al.*, 2004).

We first consider the connection between selective influence on the object processing times and the order in which those objects are completed. Above, we maintained the possibility that the order in which the processes complete affects completion time distributions in both serial and parallel systems (**SP2**). If a system has **SP2**, then a factor that influences the likelihood of an object finishing in a different order relative to other objects cannot *selectively* influence that completion time. Suppose that the factor associated with X_2 makes it faster so that for all ts $P(T_2^{\text{fast}} \leq t) \geq P(T_2^{\text{slow}} \leq t)$ with strict ordering for at least some t, where $T_2^{\text{fast}} = \lambda_2(\alpha_2^{\text{fast}}, S_2, C), T_2^{\text{slow}} = \lambda_2(\alpha_2^{\text{slow}}, S_2, C)$. Then, in a parallel system, there is a higher chance that object 2 will be faster than object 1, $P(T_2^{\text{fast}} < T_1) > P(T_2^{\text{slow}} < T_1)$, so α_2 can in general influence what position object 1 ends in. If the position in which object 1 finishes affects its completion time distribution, then T_1 with α_2^{fast} must be different from T_1 with α_2^{slow}, but this indicates a failure of selective influence. Alternatively, consider a serial system in which the probability of object 1 being processed first is selectively influenced by α_1. If the probability of object 1 being first is higher, then the probability of the other objects being first is lower. This means that α_1 affects the position of the other objects, so again, if the completion time distribution of the other objects depends on their positions, then selective influence fails.

With selective influence, we can derive a necessary condition for parallel systems that is not necessary for serial systems based solely on observing the order in which items complete. Selective influence implies marginal selectivity, so the distribution of any pair of completion times (T_i, T_j) does not vary based on factors that selectively influence objects other than i and j. In particular, orderings in a parallel system, i.e., $P\left(T_i < T_j\right)$, may be influenced by α_i and α_j but not by other factors. This implies that the probability that object 1 is completed before object 2 is the same regardless of the factors that influence the processing of object 3. Thus, in a parallel system with selective influence, $P\left(\mathcal{O}(1,2,3) \cup \mathcal{O}(1,3,2) \cup \mathcal{O}(3,1,2)\right)$ is invariant across levels of the factor associated with object 3. In contrast, in a serial process, selective influence on the completion times does not preclude the possibility that the factors influence the processing order (as long as the completion time distributions do not vary based on processing order).

We can derive another test for parallel processing using the total completion times. If the system is parallel, then the total completion time for object i is T_i, which is unaffected by the factor settings of all other objects. If the system is serial, then the total completion time for object i is the sum of all completion times for objects that complete before i and T_i. Suppose α_j influences the processing time of object j ($j \neq i$). If there is some positive probability that object j completes before object i, then the total completion time for object i will sometimes include T_j and hence be influenced by α_j.

Another thing to note about the mapping between parallel and serial systems is that the distribution over orderings in a serial system can be independent of processing time distributions and can even be unrelated to the objects to be processed. On the other hand, in a parallel system, the ordering is dependent on relatively how fast each process completes, which can certainly vary based on object properties (**SP1**, **P3**) and even on other processes (**P2**). Hence, if the order in a serial system does not depend on object properties, then it can be distinguished by examining whether the distribution of order is influenced by changes in the object properties. This leads us into the next section on distinguishing serial and parallel systems when there is an option to selectively influence a process.

Example 3.4 As a special case, there is the situation when the processing order is deterministic, e.g., a particular order has probability 1. Intuitively, this should only be possible for serial systems. However, the outlined properties do not trivially constrain parallel processes so much that they cannot have this property. Consider a three sub-process system as in Example 3.3. Thus, the stopping regions are $[1, 2], [3, 4]$, and $[5, 6]$ for each respective sub-process X_1, X_2, and X_3. Now allow the sub-processes to have the following construction:

- $X_1(t)$ follows a $unif(0, 1)$ distribution for each $t \leq 2$ and follows $unif[1, 2]$ distribution for $t > 2$.
- $X_2(t)$ follows a $unif(0, 1)$ distribution for each $t \leq 3$ and follows $unif[3, 4]$ distribution for $t > 3$.
- $X_3(t)$ follows a $unif(0, 1)$ distribution for each $t \leq 4$ and follows $unif[5, 6]$ distribution for $t > 4$.

Each X_i is being worked on from time 0 until it is complete. Also, each sub-process X_i is guaranteed (probability = 1) to finish at time $t = i + 1$. Thus, the processing order is deterministic. Note that modifying the definition of "being worked on" to require that the process could finish at any time, not just change state, would negate this example.

Finally we come to the classic survivor interaction contrast tests developed by Townsend and Nozawa (1995) and extended by Schweickert, Dzhafarov and colleagues (Schweickert *et al.*, 2000, 2012; Dzhafarov *et al.*, 2004; Zhang & Dzhafarov, 2015).

Consider first a system with two independent sub-processes that can be sped up and slowed down by selective influence manipulations. The survivor interaction contrast is defined by an interaction contrast of the survivor functions of the system completion times under two levels of the selective influence manipulation on each sub-process. In particular, suppose there is some manipulation that speeds up or slows down each process selectively,

$$T_1^H = \lambda_1\left(\alpha_1^H, S_1\right)$$
$$T_1^L = \lambda_1\left(\alpha_1^L, S_1\right)$$

such that for all t, $P\left(T_1^H > t\right) \leq P\left(T_1^L > t\right)$ with strict inequality for some t, i.e., T_1^H stochastically dominates T_1^L.

The survivor interaction contrast is given by

$$\mathrm{SIC(t)} = \left(\mathrm{S^{LL}} - \mathrm{S^{LH}}\right) - \left(\mathrm{S^{HL}} - \mathrm{S^{HH}}\right) = \left(\mathrm{F^{HL}} - \mathrm{F^{HH}}\right) - \left(\mathrm{F^{LL}} - \mathrm{F^{LH}}\right).$$

In a parallel system that stops as soon as either process has finished and independence between T_1 and T_2, the survivor function for the system completion time is given by the probability that neither of the processes have completed by time t,

$$S(t) = S_1(t)S_2(t).$$

Hence, the survivor interaction contrast can be rewritten as

$$\begin{aligned}\mathrm{SIC(t)} &= \left(S_1^L(t)S_2^L(t) - S_1^L(t)S_2^H(t)\right) - \left(S_1^H(t)S_2^L(t) - S_1^H(t)S_2^H(t)\right)\\ &= S_1^L(t)\left(S_2^L(t) - S_2^H(t)\right) - S_1^H(t)\left(S_2^L(t) - S_1^H(t)\right)\\ &= \left(S_1^L(t) - S_1^H\right)\left(S_2^L(t) - S_2^H(t)\right).\end{aligned}$$

Because of the stochastic dominance assumption, each term in the product must be positive so the product is positive and hence the survivor interaction contrast for a parallel, first-terminating process is positive for all t. Intuitively, we might expect this from the interaction contrast because the minimum of the two sub-processes should be affected more by going from L to H when the other sub-process is L compared to going from L to H when the other sub-process is already H.

If the parallel system stops only when both sub-processes complete, again assuming independence between T_1 and T_2, then the survivor function for the system completion time is given by the probability that neither sub-process is still continuing,

$$S(t) = 1 - (1 - S_1(t))(1 - S_2(t)) = S_1(t) + S_2(t) - S_1(t)S_2(t).$$

The survivor interaction contrast is given by

$$\begin{aligned}\mathrm{SIC(t)} &= \left(S_1^L(t) + S_2^L(t) - S_1^L(t)S_2^L(t) - \left(S_1^L(t) + S_2^H(t) - S_1^L(t)S_2^H(t)\right)\right)\\ &\quad - \left(S_1^H(t) + S_2^L(t) - S_1^H(t)S_2^L(t) - \left(S_1^H(t) + S_2^H(t) - S_1^H(t)S_2^H(t)\right)\right)\\ &= \left(-S_1^L(t)S_2^L(t) + S_1^L(t)S_2^H(t)\right) - \left(-S_1^H(t)S_2^L(t) + S_1^H(t)S_2^H(t)\right)\\ &= S_1^L(t)\left(-S_2^L(t) + S_2^H(t)\right) - S_1^H(t)\left(-S_2^L(t) + S_2^H(t)\right)\\ &= \left(S_1^L(t) - S_1^H(t)\right)\left(S_2^H(t) - S_2^L(t)\right).\end{aligned}$$

Again, relying on stochastic dominance from the selective influence manipulation, the first term in the final product is positive for all t while the second term is negative for all t, indicating that the survivor interaction contrast for the parallel, exhaustive condition is negative for all t. The intuition for this result mirrors the intuition for the parallel, first-terminating model. If the slowest of the processes determines the system completion time, then going from L to H when the other process is already L will have less of an effect than going from L to H when the other process is H.

In a serial system that stops as soon as either sub-process is finished that has some probability p of a particular sub-process being processed first (and assuming that the selective influence manipulations on the sub-process completion times do not affect p), the completion time distribution is

$$F(t) = pF_1(t) + (1-p)F_2(t).$$

Hence, the survivor interaction contrast is

$$\begin{aligned}\text{SIC(t)} &= \left(pF_1^H(t) + (1-p)F_2^L(t) - pF_1^H(t) - (1-p)F_2^H(t)\right) \\ &\quad - \left(pF_1^L(t) + (1-p)F_2^L(t) - pF_1^L(t) - (1-p)F_2^H(t)\right) \\ &= \left((1-p)F_2^L(t) - (1-p)F_2^H(t)\right) - \left((1-p)F_2^L(t)\right. \\ &\quad \left.-(1-p)F_2^H(t)\right) = 0.\end{aligned}$$

Finally, for the serial model which requires both sub-processes to finish, the distribution is the convolution of the completion time density of one sub-process with the completion time distribution of the other sub-process,

$$S(t) = \int_0^\infty f_1(s)S_2(t-s)\,ds = f_1 * S_2.$$

This means the survivor interaction contrast is given by

$$\begin{aligned}\text{SIC(t)} &= \left(\int_0^\infty f_1^L(s)S_2^L(t-s)\,ds - \int_0^\infty f_1^L(s)S_2^H(t-s)\,ds\right) \\ &\quad - \left(\int_0^\infty f_1^H(s)S_2^L(t-s)\,ds - \int_0^\infty f_1^H(s)S_2^H(t-s)\,ds\right) \\ &= \int_0^\infty \left(f_1^L(s)\left(S_2^L(t-s) - S_2^H(t-s)\right)\right. \\ &\quad \left. - f_1^H(s)\left(S_2^L(t-s) - S_2^H(t-s)\right)\right)\,ds \\ &= \int_0^\infty \left(f_1^L(s) - f_1^H(s)\right)\left(S_2^L(t-s) - S_2^H(t-s)\right)\,ds.\end{aligned}$$

Due to the stochastic dominance from the selective influence manipulation, the difference of survivor functions in the product will always be positive, so the difference of densities determines the sign of the integrand. Because of the dominance on the first sub-process, $f_1^L(t)$ must be smaller than $f_1^H(t)$ for some range of time $(0, t^*)$, so the integrand, and hence the SIC, is negative for some initial range of times. To verify that the SIC is also positive for some range of time, it is sufficient to show that the integrated SIC is 0. Because of the relationship (for positive random variables), $\int_0^\infty S_X(t)\,dt = E[X]$, the integrated survivor interaction contrast is the interaction contrast of the means,

$$\int_0^\infty \mathrm{SIC(t)}\,\mathrm{dt} = (\mathrm{E}\left[\mathrm{T}^{\mathrm{LL}}\right] - \mathrm{E}\left[\mathrm{T}^{\mathrm{LH}}\right]) - (\mathrm{E}\left[\mathrm{T}^{\mathrm{HL}}\right] - \mathrm{E}\left[\mathrm{T}^{\mathrm{HH}}\right]).$$

Because the expectation is a linear operator, $E[T] = E[T_1] + E[T_2]$, so

$$\begin{aligned}\int_0^\infty \mathrm{SIC(t)}\,\mathrm{dt} = &\left(E\left[T_1^L\right] + E\left[T_1^L\right] - E\left[T_1^L\right] - E\left[T_1^H\right]\right) \\ &- \left(E\left[T_1^H\right] + E\left[T_1^L\right] - E\left[T_1^H\right] - E\left[T_1^H\right]\right) = 0.\end{aligned}$$

In more recent work, Yang, Fifić, and Townsend (2012) demonstrated that, for most cases, there is only a single zero-crossing for the serial-exhaustive survivor interaction contrast. For our current purposes, it is sufficient that we have demonstrated that using the survivor interaction contrast with selective influence manipulations, serial and parallel processes can be distinguished.

Although these proofs all depended on independence between the completion time distributions of the sub-processes, as long as selective influence holds, the same conclusions hold. By the selective influence definition above, the completion time of one sub-process is *conditionally* independent of the other sub-process completion time given C. Hence, if we replace the densities, distributions and survivor functions above with the conditional versions, then the derivations hold for the dependent, but selectively influenced, processes. To extend from there to the unconditional distributions, note that the conditional holds when conditioned on any C so the derivations above are true when integrated across all C, i.e., when C is marginalized.

3.6 Discussion and Conclusions

In this chapter, it has been our goal to formalize the concepts of parallel and serial systems in cognitive modeling. Based on this foundation, we have endeavored to enumerate the contexts in which these systems are discriminable and those contexts in which they can perfectly mimic one another.

This chapter is by no means the first treatise on this topic. Studies concerning the identifiability of serial and parallel systems have appeared since the early days of cognitive psychology. We have attempted to draw attention to the landmarks in the development of this important topic, but because our goal was to illuminate the issues of serial and parallel identifiability themselves rather than their history, we have necessarily omitted direct reference to important works in this domain.

What we have done is to begin with a modification of the informal statement of the shared and distinctive properties of parallel and serial systems posited by Townsend and Ashby (1983). From there, we gave a brief overview of the necessary measure-theoretic probability theory to introduce the concepts of filtrations and adapted stochastic processes. We then used the formalism of an adapted stochastic process as the foundation of our definitions of parallel and serial systems. The use of adapted stochastic processes has allowed us to formalize the

notion that these systems should maintain causality, particularly the constraint that the future should not influence the present. These formal systems were then used to explore the necessary (and sometimes sufficient) conditions under which serial and parallel systems may be discriminated. In many cases, serial and parallel structures are only discriminable under selective influence manipulations, including some of the most-used theorems that we covered, those concerning the survivor interaction contrast.

As part of our development, we have also highlighted many instances in which parallel and serial systems are *not* discriminable. This may serve as a cautionary note to empirical researchers not to quickly label cognitive systems as parallel or serial without considering these possible mimicking issues.

On a more positive note, we hope that these theoretical developments may serve as a foundation for future research that might result in more, and more powerful, tests for empirically discriminating between these two classes of systems.

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 *Oxford library of psychology*. *The Oxford handbook of computational and mathematical psychology* (pp. 63–98). Oxford: Oxford University Press.

Billingsley, P. (1995). *Probability and measure*, third edition. New York, NY: Wiley.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Dzhafarov, E. N. (2003). Selective influence through conditional independence. *Psychometrika*, *68*, 7–26.

Dzhafarov, E. N., & Gluhovsky, I. (2006). Notes on selective influence, probabilistic causality, and probabilistic dimensionality. *Journal of Mathematical Psychology*, *50*, 390–401.

Dzhafarov, E. N., & Kujala, J. V. (2010). The joint distribution criterion and the distance tests for selective probabilistic causality. *Frontiers in Psychology*, *1*, 151.

Dzhafarov, E. N., Schweickert, R., & Sung, K. (2004). Mental architectures with selectively influenced but stochastically interdependent components. *Journal of Mathematical Psychology*, *48*, 51–64.

Houpt, J. W., & Townsend, J. T. (2011). An extension of SIC predictions to the Wiener coactive model. *Journal of Mathematical Psychology*, *55*, 267–270.

Kujala, J. V., & Dzhafarov, E. N. (2008). Testing for selectivity in the dependence of random variables on external factors. *Journal of Mathematical Psychology*, 52, 128–144.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

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* (Vol. 4). Singapore: World Scientific.

Schweickert, R., Giorgini, M., & Dzhafarov, E. (2000). Selective influence and response time cumulative distribution functions in serial–parallel task networks. *Journal of Mathematical Psychology*, *44*, 504–535.

Snodgrass, J. G., & Townsend, J. T. (1980). Comparing parallel and serial models: Theory and implementation. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 330–354.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, *4*, 421–457.

Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (ed.) *Attention and performance*, vol. 4 (pp. 629–685). New York: Academic Press.

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. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology*, *14*, 219–238.

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

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, *1*, 46–54.

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

Townsend, J. T., & Evans, R. (1983). A systems approach to parallel-serial testability and visual feature processing. In H. G. Geissler (ed.), *Modern issues in perception* (pp. 166–189). Berlin: VEB Deutscher Verlag der Wissenschaften.

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*, 321–360.

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. D. (1994). Stochastic dependencies in parallel and serial models: Effects on systems factorial interactions. *Journal of Mathematical Psychology*, *38*, 1–24.

Van Zandt, T., & Townsend, J. T. (2012). Designs for and analyses of response time experiments. In T. K. Little (ed.), *The Oxford handbook of quantitative methods* (pp. 260–285). New York, NY: Oxford University Press.

Yang, H., Fifić, M., & Townsend, J. T. (2012). Survivor interaction contrast wiggle predictions of parallel and serial models for an arbitrary number of subprocesses. *Journal of Mathematical Psychology*, *48*, 21–32.

Zhang, R., & Dzhafarov, E. (2015). Noncontextuality with marginal selectivity in reconstructing mental architectures. *Frontiers in Psychology*, *6*(735). Available from www.frontiersin.org/cognition/10.3389/fpsyg.2015.00735/abstract