CHAPTER 11

Uncovering Mental Architecture and Related Mechanisms in Elementary Human Perception, Cognition, and Action

JAMES T. TOWNSEND, MICHAEL J. WENGER, AND JOSEPH W. HOUPT

INTRODUCTION: BRIEF HISTORY AND GENERAL CONCEPTION

Psychology, the study of the mind, takes for itself behavioral as well as surgical and physiological-recording sets of tools. The behavioral and physiological data can be interpreted, and sometimes predicted, in verbal as well as mathematical languages.

Our approach lies within the confines of mathematical theory and methodology and concentrates on behavioral data, although some inroads and prospects for fusion with physiological techniques will be mentioned. To the extent that psychology and cognitive science wish to go beyond an extreme behavioristic catalog of stimulus-response correlations, they are what theoretically inclined engineers would call *black box sciences* (e.g., see Booth, 1967; Townsend, 1984). The very essence of black box sciences is to utilize input-output sequence data to uncover the inner subprocesses and their interactions that can generate that data.[1]

The history of black box psychology most concretely starts in the 19th-century laboratories of pioneering scientists such as F. C. Donders and W. Wundt. These men in particular were perhaps the very first to manipulate the components of the stimuli and study response times (RTs) in order to draw inferences about the underlying mechanisms involved in elementary perceptual, cognitive, and action processes. There are many other roots of embryonic psychological science that complete the background for the current chapter, including ingenious observations of that other major dependent (observable) variable of psychology: patterns of accuracy. However, the so-called *complication experiments* carried out by Donders, Wundt, and their students, colleagues, and intellectual descendants constitute the true historical genesis of our methodology.

The complication experiment provided a task, such as responding as soon as possible to the onset of a signal. Then, in separate experimental trials, the researcher would complicate the task by adding another subtask, such as requiring a discrimination between two separate signals. A critical postulate was that the extra subtask was placed in sequence with the already attendant subtask

---

[1] Naturally, the strategies required to identify the mechanisms, possibly including germane physics or chemistry, must be drastically different from those involved in, say, electronic circuits or automata in general.

(Woodworth, 1938, Chapter 14). In more contemporary terminology, this postulate can be understood in terms of the assumptions of *pure insertion* of the new subtask plus *serial processing* (e.g., Ashby & Townsend, 1980; Sternberg, 1966, 1969; for a thorough discussion of many such issues in psychology and cognitive science see Lachman, Lachman, & Butterfield, 1979). By subtracting the mean RT of the simpler task from that of the more complicated task, an estimate of the mean time of the inserted subtask could be calculated.

In fact, the pair of opposing hypotheses— serial versus parallel processing, or the question of *mental architecture*—though hinted at in research dating back to around 1950, became a primary, essential question in the growing cognitive science of the 1960s and beyond (e.g., Egeth, 1966; Sternberg, 1966). It is important to note that *mental architecture* is not meant to imply a fixed, unchangeable structure. For example, there is increasing evidence that some individuals may be able to alter the processing mode from serial to parallel or vice versa (e.g., Townsend & Fifić, 2004; Yang, Little, & Hsu, 2014).

Casting the aim of psychological theory and methodology in terms of a search for mechanisms in a very complex black box might bring along with it the worry that, in the absence of an exact and complete dissection of the black box's innards, different mechanisms and different interactions of similar mechanisms might be capable of mimicking one another. For instance, in discrete automata theory, given even an infinite series of input-output observations, there will still be a set of systems, each of which could generate that observational corpus. These would be distinct from the "falsified" class of systems but would all be capable of producing the observed data, and thus of mimicking each other (again, see how this works out in the well-studied domain of finite state automata: Booth, 1967).

The issue of mimicry has always been present, but only rarely explicitly recognized or studied in a rigorous fashion. In modern psychology, perhaps its first notable appearance was in the analysis of mimicking of Markov models of memory (Greeno & Steiner, 1964). The analysis of mimicry in stochastic accumulator models of perception and decision has come more recently (e.g., Dzhafarov, 1993; Jones & Dzhafarov, 2014; Khodadadi & Townsend, 2015; Link & Heath, 1975; Ratcliff, 1978). As will become clear, our consideration of theoretical mimicry raises the question of the experimental methodologies needed to potentially resolve questions of mimicry. Throughout most of the history of psychology, theory and methodology have been almost entirely segregated, with theory being largely verbal and methodology being concentrated in statistics. Our approach views theory and methodology as Siamese twins, with the cogent application of each requiring consideration of the other.

As already intimated, our approach for over 40 years has been mathematical in nature. Furthermore, much of our work has been what we refer to as *metatheoretical*, in the sense that our major aim has been to study simultaneously large classes of mathematical models in relation to well-specified types of experimental designs. The aim is to discover which classes of models mimic one another and which models can be rigorously tested against one another in these paradigms. Of course, the paradigms and the model classes are optimally developed in close tandem with one another.[2] One of the strengths of the metatheoretical approach is that it is usually possible to render the classes

---

[2]This approach can also be referred to as *qualitative modeling*. However, that terminology can be mistaken as referring to purely verbal inquiries.

of assayed model in terms of general probability distributions and the assays themselves in terms of nonparametric statistics and tests.

Allen Newell, a pioneer and luminary in the development of cognitive science, implicitly brought in the mimicking dilemma in his widely cited article "You Can't Play 20 Questions with Nature and Win..." (1973). Newell called into question the ability of scientists to effectively test between dichotomous properties and issues in cognition. In contrast, the method of *strong inference* (Platt, 1964), which recommends that scientists pose a series of ever more specific alternative hypotheses to test against one another in order to avoid confirmation bias, does just this. Interestingly, Greenwald (2012) cogently promotes the indissoluble and mutually supportive nature of method and theory. In "In a Revisit to Newell's 20 Questions," (2012), Greenwald finds evidence in our approach for a solution to Sternberg's (1966) issue of parallel versus serial processing.

The metatheoretical approach to date has preponderantly relied on the two strongest dependent variables, RT and patterns of response probabilities. It is usually accepted that RT lies on a ratio scale while probabilities lie on an absolute scale (e.g., see Krantz, Luce, Suppes, & Tversky, 1971; Townsend, 1992; Townsend & Ashby, 1984). In addition, it turns out that the nonparametric assays employed in our metatheoretical approach are invariant up to monotonic transformation, meaning that measurement scales that are only ordinal can be implemented without violating principles of foundational measurement (again, see Krantz et al., 1971, and Chapter 12 of Townsend & Ashby, 1983).

The more frequent style of modeling in psychology and cognitive science has been to propose a detailed, parameterized model for certain behaviors and tasks (see also Chapter 9 in this volume). For example, Ratcliff (1978) proposed a model for memory

search based on the well-known Wiener diffusion process. That model has since gained wide recognition and application. Ratcliff's Wiener process is continuous in both time and state. Even earlier, Link and colleagues (e.g., Link & Heath, 1975) invented a mathematically related approach, also based on continuous state but using discrete time: a sequential random walk model later referred to as *relative judgment theory*. Interestingly, they emphasized broad assessments of a more qualitative nature than the more typical model fits. Both approaches are valuable and complementary. Thus, one salutary tactic is to use the metatheoretical strategy to discern the broad characteristics of the mental operations and then to bring to bear specific process models that obey the dictates of the broad elements discovered through the former approach (e.g., see Eidels, Donkin, Brown, & Heathcote, 2010).

This chapter provides the most up-to-date and general treatment of our theoretical and methodological philosophy and science. Limitations and pertinent references to related material are addressed in the conclusion. A somewhat more mathematical treatment and one that expands on some of the topics necessarily abbreviated here is available in Algom, Eidels, Hawkins, Jefferson, and Townsend (2015).

MAJOR CHARACTERISTICS OF ELEMENTARY COGNITIVE SYSTEMS AND THEOREMS OF MIMICRY: THE EVENT SPACE BASIS

According to modern probability theory (dating to Kolmogorov, 1950), a probability space can be informally defined as follows: It is made up of a family of sets—the events in question—together with the set operations on them, plus the imposition of a probability

measure. The measure on the entire set of points from which the sets are taken must be 1, and the empty set has probability 0.[3] The primary elements of the probability spaces used for the consideration of mental architecture have been the item-processing times combined with the order of completion (e.g., Townsend, 1972; Townsend, 1976a, 1976b). Formal representation of this type of event space was reported by Vorberg (1977) at a meeting of the Society for Mathematical Psychology, and followed in Townsend and Ashby (1983, Chapter 14). The remainder of this section will adhere to an examination of this type of space, although the discussion will remain in an informal mode. Subsequently we will enrich the spaces for serial and parallel models, but again, will keep to a vernacular modus operandi.

Because the Sternberg paradigm (Sternberg, 1966) has been one of the most popular and replicated paradigms in the history of human experimental psychology, we will utilize it to illustrate the critical issues in human information processing. In this paradigm, the experimenter presents a modest number of discrete items; for instance, letters or numbers. The number is typically varied—say, from one to six—with no repetitions. After a very brief interval, a probe item is presented and the observer's task is to indicate whether or not that probe was contained in the original set of items. The observer's RTs are recorded and the RT means graphed as a function of the number of items, with *yes* versus *no* curves acting as a parameter in the plots and statistical analyses. An analogous procedure can be carried out for visual search, with the probe presented first, followed by a varied number of items in a brief visual display (e.g., Atkinson,

Holmgren, & Juola, 1969). The observer is then required to respond *yes* versus *no* depending on whether the target item was found in the visual list or not.

Now, suppose that pure insertion holds for each of the n items so that memory search is serial; that is, mean RT is an additive function of n, plus a *residual time* assumed to be occupied by earlier sensory processes, ancillary cognitive terms, and late-term response processes, which are postulated to be independent and invariant across values of n. If in addition the purely inserted times are identically distributed, then the mean RT functions will be linear. Stochastic independence of the successive stage times (*across-stage independence*) is typically assumed too, but this constraint is not strictly necessary. In Sternberg's (1966) data, the mean RT functions were indeed linear, thus being consistent with the hypotheses of serial processing, pure insertion, and identically distributed increments.

Furthermore, Sternberg asked if the mean RT functions for *yes* and *no* responses possessed the same slope. Sternberg reasoned that *no* responses necessitated that all memory items be processed, a so-called *exhaustive stopping rule*. However, the observer could stop on *yes* trials as soon as the probe was located in the memory list, a so-called "*self-terminating*" *stopping rule*. Under the other assumptions about seriality and identical process-time distributions, the *yes* curve should thus have one half of the slope of the *no* curve. Instead, Sternberg found equal slopes: Mean RT as a function of n was linear for both *yes* and *no* responses. And, the slopes were deemed to be equal, suggesting exhaustive processing for both types of responses. Over the years, the findings of the linearity and equal-slopes result has often, but not universally, been replicated. Equal slopes are somewhat more rare in the case of visual search (e.g., see Schneider &

---

[3]A succinct, but elegant, introduction to modern probability theory can be found in Chung (2001). Billingsley (1986) is more complete with a nice survey of the topic.

Shiffrin, 1977; Townsend & Roos, 1973; also see Nosofsky, Little, Donkin, & Fifić, 2011).

Another indicant of the stopping rule must be mentioned, and that is the phenomenon of *position effect*. If mean RTs are shorter if the probe appears in some positions more than others, this could be, and usually is, interpreted to mean that self-termination is in effect and that some paths of serial processing are more often taken than others. This issue will be discussed further below (also see Townsend & Ashby, 1983, Chapter 7). We begin our formal description of the issues, beginning with a set of definitions.

Definition 1: A stochastic model of a system for processing positions *a* and *b* is *standard serial* with an exhaustive processing rule if and only if the density functions of the serial processing times satisfy

$$f[t_{a1}, t_{b2}; (a,b)] = pf(t_{a1})f(t_{b2})$$

and

$$f[t_{b1}, t_{a2}; (b,a)] = (1-p)f(t_{b1})f(t_{a2}).$$

First, note that Definition 1 includes the assumption of across-stage independence. Definition 1 is given for $n = 2$. If we extend it to arbitrary *n*, we directly acquire the increasing straight-line predictions for mean RTs from Sternberg (1966, 1975). Of course, if processing is self-terminating

on probe-present trials, then we expect the mean RT to possess a slope one half of the probe-absent trials. Sternberg's experimental results, along with many replications by others, certainly appear to be very persuasive: The logic is impeccable and the findings oft replicated with good statistical power. But then how can mimicking occur in the Sternberg task? The most antipodal type of processing that could occur is *parallel processing*. Such processing, in contrast to serial processing, postulates that all items are compared with the probe at once, thus seemingly predicting much shorter mean RTs as *n* is varied. Consequently, we need a definition for parallel models.

Definition 2: A stochastic model of a system for processing positions *a* and *b* is parallel with an exhaustive processing rule if and only if the density functions on the parallel processing times satisfy

$$g_{a1,b2}[\tau_{a1}, \tau_{b2}; (a,b)] \text{ with } \tau_{b2} > \tau_{a1}$$

and

$$g_{a2,b1}[\tau_{a2}, \tau_{b1}; (b,a)] \text{ with } \tau_{b1} < \tau_{a2}.$$

It can be seen that the parallel processing times are, in contrast to serial processing, overlapping, as indicated in Figure 11.1. A structural relationship between the *ts* and the τs is that $\tau_{r1} = t_{r1}$ for $r = a$, *b* and
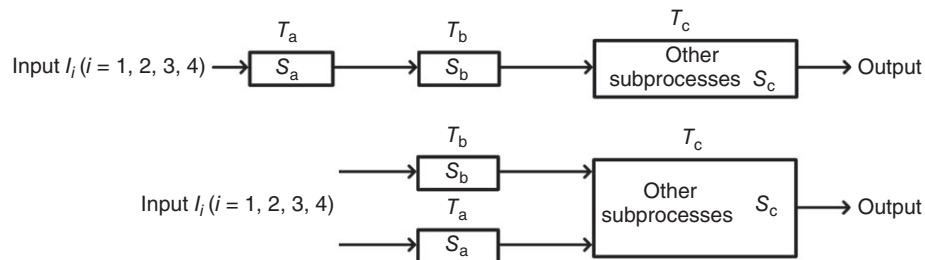

Figure 11.1   Schematic of serial (upper) and parallel (lower) processing systems. The subprocesses S_a, S_b, and S_c are each selectively influenced by a unique factor (or set of factors), T_a, T_b, and T_c, respectively.

$\tau_{b2} = ta_1 + tb_2$ and $\tau_{a2} = t_{b1} + t_{a2}$. An equivalent statement is that $t_{b2} = \tau_{b2} - \tau_{a1}$ and $t_{a2} = \tau_{a2} - \tau_{b1}$. See Townsend and Ashby (1983, Chapter 1) for details.

A useful special case arises when the parallel processing times are *within-stage independent*, which indicates that between successive completions the parallel channels are operating in a probabilistically independent fashion. It is easiest to pose this condition if we use the above interpretation in terms of the *t*'s, which are, after all, the *inter-completion times*. First, let $G(t) = P(T > t)$ and $G^*(t) = 1 - G(t)$ for any particular channel. This leads us to our third definition.

Definition 3: A stochastic model of a system for processing positions *a* and *b* is parallel, which satisfies *within-stage independence* if and only if the density functions on the parallel processing times during the first stage satisfy

$$g_{a1,b2}[t_{a1}, t_{b2}; (a,b)]$$
$$= g_{a1}(t_{a1})\, G^*_{b1}(t_{a1}) G_{b2}(t_{b2}|t_{a1})$$

and

$$g_{a2,b1}[\tau_{a2}, \tau_{b1}; (b,a)]$$
$$= g_{b1}(t_{b1})G^*_{a1}(t_{b1})\, g_{a2}(t_{a2}|t_{b1}).$$

Note that within-stage independence and across-stage independence are logically independent notions. Finally, the counterpart to the standard serial model is the standard parallel model, which possesses independence of the actual parallel processing times τ_a and τ_b. This model is described in our fourth definition.

Definition 4: A stochastic model of a system for processing positions *a* and *b* is *standard parallel* with an exhaustive processing rule if and only if the density functions on the parallel processing times satisfy

$$g_{a1,b2}[\tau_{a1}, \tau_{b2}; (a,b)] = g_a(\tau_{a1})g_b(\tau_{b2})$$

with $\tau_{b2} > \tau_{a1}$ and

$$g_{a2,b1}[\tau_{a2}, \tau_{b1}; (a,b)] = g_a(\tau_{a2})g_b(\tau_{a2})$$

with $\tau_{b1} < \tau_{a2}$.

It is important to be aware that independence of parallel processing times and across-stage independence are separate and logically independent concepts. Any model, parallel or serial, could satisfy either or not (Townsend & Ashby, 1983, Chapter 4).

Now, suppose that processing is indeed parallel with stochastic independence (thus satisfying Definition 4), and that every item is operated on at the same speed (i.e., identically distributed processing times), and that these distributions are invariant with *n*, implying *unlimited capacity*.

Then, Townsend and Ashby (1983, Chapter 4) showed that if processing is exhaustive, mean RTs increase in a curvilinear manner, with each added item adding a steadily decreasing increment to the total completion time of all the items. If processing is self-terminating, the mean RT function will be flat. Sternberg's (1966) results were not well fit by such curves, therefore falsifying the standard type of parallel model with identical processing-time random variables. Even if the assumption of identically distributed processing times for the distinct items is relaxed, such models will still not mimic the standard serial predictions.[4]

However, if the assumption of unlimited capacity is dropped so that as *n* increases, the item-processing times slow down, then parallel models can easily mimic the linear increasing-mean RT functions. In fact, suppose that as *n* increases, a fixed amount of capacity must be spread across all the memory items (or in visual search, the presented items). Then the first item to be processed will

---

[4]Note that if distinct positions of items possess distinct processing-time distributions, and if processing is self-terminating, position effects can be readily predicted by these parallel models.

take up the same time as the first serial item to be processed. Moreover, if the capacity devoted to the completed item is reallocated to all the unfinished items, then each succeeding item to be processed will add the same amount of time to the overall completion time. Given any set of serial distributions, such parallel models can be mathematically identical, at both the level of the means and the level of the distributions (Townsend, 1969, 1971, 1972, 1974). Interestingly, if only the mean RTs are considered, rather than the entire distributions, the assumption of reallocation of capacity can be eliminated. Chapter 4 in Townsend and Ashby (1983) shows how architecture, stopping rule, and workload capacity combine to produce a variety of mean RT curves as functions of n.

There isn't space to delve into technical issues of parallel-serial mimicry in detail, much less the complete history. However, in our opinion, mimicry and testability are akin to the yin and yang of Buddhist philosophy: The universe of opposing models cannot be fully understood unless both are objects of deliberation. Consequently, we need definitions that directly address the issue. Let T_{ij} stand for the random variable for item or position $i = a, b$ and $j = 1, 2$.

Definition 5: The *fundamental functional equation of mimicry* for $n = 2$ is

$$p[f_{a1}(t_{a1})f_{b2}(t_{b2}|t_{a1})]$$
$$= g_{a1,b2}[\tau_{a1}, \tau_{b2};\ (a, b)]$$
$$= g_{a1,b2}(t_{a1}, t_{a1} + t_{b2})$$
$$= g_{a1}(t_{a1}, T_{b1} > t_{a1})g_{b2}(t_{b2}|t_{a1})$$

and

$$(1 - p)[f_{b1}(t_{b1})f_{a2}(t_{a2}|t_{b1})]$$
$$= g_{a2,b1}[\tau_{a2}, \tau_{b1};(b, a)]$$
$$= g_{a2,b1}(t_{b1} + t_{a2}, t_{b1})$$
$$= g_{b1}(t_{b1}, Ta_1 > t_{b1})g_{a2}(t_{a2}|t_{b1})$$

for all $t_{ai}, t_{bj} > 0, i, j = 1,\ 2$.

If within-stage independence is assumed on the part of the parallel models, then we achieve the next definition.

Definition 6 (Townsend, 1976a): If within-stage independence holds for the parallel class of models then the *fundamental equation of mimicry* for $n = 2$ is

$$\text{Serial} \qquad\qquad \text{Parallel}$$
$$p[f_{a1}(t_{a1})f_{b2}(t_{b2}|t_{a1})] = g_{a1}(t_{a1})G^*_{b1}(t_{a1})$$

and

$$(1 - p)[f_{b1}(t_{b1})f_{a2}(t_{a2}|t_{b1})] = g_{b1}(t_{b1})G^*_{a1}(t_{a1})$$

for all $t_{ai}, t_{bj} > 0, i, j = 1,\ 2$.

For every specification of a parallel model (the right-hand side) there exists a serial model; that is, values of p and the serial density functions that can perfectly mimic the parallel model. However, when the equation is solved for the parallel-model functions in terms of the serial functions, the result may not result in a distribution whose cumulative function properly goes to 1. Townsend (1976a) interpreted such solutions as violating the definition of parallel processing. However, Vorberg (1977) pointed out that an alternative is that the system is still parallel, but that on some trials, a channel fails to complete processing.

The instantiation of within-stage independence might seem like a stringent condition. However, it turns out that any joint distribution not satisfying within-stage independence can be transformed into within-stage independent distributions (Rao, 1992). So far, our emphasis has been on how these polar-opposite types of architecture can act so much like one another. In the succeeding sections, we focus on tactics for testing parallel versus serial processing against one another.

STRONG EXPERIMENTAL TESTS ON RESPONSE TIMES I: EVENT SPACE EXPANSION

The event spaces employed in the preceding section are too coarse for a complete treatment of mental architectures. However, we have seen that even with the former, coarse description, the serial models can be more general than the parallel if either (a) we remain within the exponential class of models (e.g., Townsend, 1972), or (b) we demand that the parallel processing-time distributions are nondefective (Townsend, 1976a).

Here we can expand the form of the probability space to take into explicit consideration what an item is made of. For instance, one might posit that an item is constituted of a finite number of features. The number of features completed by a certain time t then becomes the nucleus of a state space, which permits the assignment of a probability measure.

Many popular models of human perception, cognition, and action contain a finer such level description of processing, one including some type of state space that goes beyond simple *uncompleted* versus *completed*. Familiar examples are Poisson counting processes (e.g., Smith & Van Zandt, 2000), random walks (e.g., Link & Heath, 1975), and diffusion models (e.g., Busemeyer & Townsend, 1993; Ratcliff, 1978). Each of these specifies a state space representing degree of processing. For instance, a Poisson counting model might represent an item's features, with the state of processing referring to the number of completed features. In contrast, the state space for a diffusion or random walk process lies on a continuum. For instance, the typical state in such a space is a real number.

Finer grained types of spaces were presaged in certain earlier papers. Most models, ours included, have typically left out details of the psychological processes being studied. For instance, though memory and display search patently demand some type of matching process, most models leave that out of the picture. However, Townsend and colleagues (1976b; Townsend & Evans, 1983) construct stochastic processes for search in a way that the probability event spaces are indeed founded on such pattern matching. Houpt, Townsend, and Jefferson (2017) present a more detailed and rigorous foundation of the underlying and richer measure space for parallel and serial models than is possible to describe here.

This enrichment unveils ways in which parallel systems are more general than the serial systems. The enriched event spaces allow for *mismatches* (also called *negative matches*) to be associated with distributions that differ from matches (*positive matches*). A fact in search and same–different paradigms for decades has been that, modally, positive matches are faster than negative matches. This small and reasonable expansion implies that intercompletion times can depend on all of the unfinished items and not only the previous processing time and order of processing, as in the usual models that ignore positive versus negative matches (e.g., Bamber, 1969; Krueger, 1978). The associated diversity of parallel and serial models led to the design of strong tests of architecture. This type of design has been dubbed the *parallel-serial tester* (or PST; see Snodgrass & Townsend, 1980; Townsend & Ashby, 1983, Chapter 13, Chapter 15), as shown in Table 11.1.

The basic PST design requires participants to determine whether one or two probe stimuli are targets, where the target is either an item in memory, a simultaneously displayed item, or an item revealed after the probes. In one condition (PST-*loc*) the target probe is always present, and participants are asked to respond affirmatively if the target probe is in a specific spatial or temporal

Table 11.1   The Parallel-Serial Tester

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

position and negatively otherwise. With $+$ indicating that the time for positive matches to the target and $-$ indicating the time for negative matches to the target, the mean RT from a serial self-terminating system would be

$$p\mathrm{E}[T_{a+}] + (1 - p)\mathrm{E}[T_{b-}] \qquad \text{target at } a$$
$$p\mathrm{E}[T_{a-}] + (1 - p)\mathrm{E}[T_{b+}] \qquad \text{target at } b.$$

In a parallel, self-terminating system, the mean RTs would be

$$\int_0^\infty G_{a+}^*(t)G_{b-}^*(t)\,dt \qquad \text{target at } a$$
$$\int_0^\infty G_{a-}^*(t)G_{b+}^*(t)\,dt \qquad \text{target at } b.$$

In the second condition (PST-*and*), either, both, or neither of the probes may be targets and the participant is asked to respond affirmatively only if *both* probes are targets. The serial, self-terminating predictions are

$$p(\mathrm{E}[T_{a1+}] + \mathrm{E}[T_{b2+}]) \qquad \text{when both are}$$
$$+ (1 - p)(\mathrm{E}[T_{a2+}] \qquad \text{targets,}$$
$$+ \mathrm{E}[T_{b1+}])$$

$$p(\mathrm{E}[T_{a1+}] + \mathrm{E}[T_{b2-}]) \qquad \text{when only}$$
$$+ (1 - p)\mathrm{E}[T_{b1-}] \qquad \text{target at } a,$$

$$p\mathrm{E}[T_{a1-}] + (1 - p)(\mathrm{E}[T_{a2-}] \qquad \text{when only}$$
$$+ \mathrm{E}[T_{b1+}]) \qquad \text{target at } b,$$

$$p\mathrm{E}[T_{a1-}] + (1 - p)\mathrm{E}[T_{b1-}] \qquad \text{when neither}$$
$$\text{are targets.}$$

With the indicator function given by I(X) equal to 1 if X true and equal to zero otherwise, the parallel, self-terminating model predictions are

$$\int_0^\infty G_{a1+}^*(t)G_{b1+}^*(t)\,dt \qquad \text{when both are}$$
$$+ \mathrm{E}[T_{b2+}\mathrm{I}(T_{a1+} < T_{b1+})] \qquad \text{targets,}$$
$$+ \mathrm{E}[T_{a2+}\mathrm{I}(T_{b1+} < T_{a1+})]$$

$$\int_0^\infty G_{a1+}^*(t)G_{b1-}^*(t)\,dt \qquad \text{when only}$$
$$+ \mathrm{E}[T_{b2-}\mathrm{I}(T_{a1+} < T_{b1-})] \qquad \text{target at } a,$$

$$\int_0^\infty G_{a1-}^*(t)G_{b1+}^*(t)\,dt \qquad \text{when only}$$
$$+ \mathrm{E}[T_{b2-}\mathrm{I}(T_{b1+} < T_{a1-})] \qquad \text{target at } b,$$

$$\int_0^\infty G_{a1-}^*(t)G_{b1-}^*(t)\,dt \qquad \text{when neither}$$
$$\text{are targets.}$$

In the final condition (PST-*or*), the same set of trial types as the second condition is used, but the participant is asked to respond affirmatively if *either* of the probes are targets. The serial, self-terminating model predicts

$$p\mathrm{E}[T_{a1+}] + (1 - p)\mathrm{E}[T_{b1+}] \qquad \text{when both are}$$
$$\text{targets,}$$

$$p\mathrm{E}[T_{a1+}] + (1 - p)(\mathrm{E}[T_{a2+}] \qquad \text{when only}$$
$$+ \mathrm{E}[T_{b1-}]) \qquad \text{target at } a,$$

$$p(\mathrm{E}[T_{a1-}] + \mathrm{E}[T_{b2+}]) \qquad \text{when only}$$
$$+ (1 - p)\mathrm{E}[T_{b1+}] \qquad \text{target at } b,$$

$$p(\mathrm{E}[T_{a1-}] + \mathrm{E}[T_{b2-}]) \qquad \text{when neither}$$
$$+ (1 - p) \times (\mathrm{E}[T_{a2-}] \qquad \text{are targets.}$$
$$+ \mathrm{E}[T_{b1-}])$$

The parallel, self-terminating model predicts

$$\int_0^\infty G_{a1+}^*(t)G_{b1+}^*(t)\,dt \qquad \text{when both are}$$
$$\text{targets,}$$

$$\int_0^\infty G_{a1+}^*(t)G_{b1-}^*(t)\,dt \qquad \text{when only}$$
$$+ \mathrm{E}[T_{a2+}\mathrm{I}(T_{b1-} < T_{b1+})] \qquad \text{target at } a,$$

$$\int_0^\infty G_{a1-}^*(t)G_{b1+}^*(t)dt$$
$$+ \, \mathrm{E}[T_{b2+}\mathrm{I}(T_{a1-} < T_{b1+})]$$

when only target at b,

$$\int_0^\infty G_{a1-}^*(t)G_{b1-}^*(t)dt$$
$$+ \, \mathrm{E}[T_{b2-}\mathrm{I}(T_{a1-} < T_{b1-})]$$
$$+ \, \mathrm{E}[T_{a2-}\mathrm{I}(T_{b1-} < T_{a1-})]$$

when neither are targets.

That is, RTs to a present target at a, T_{a1+}, are modeled with a different random variable than RTs to a mismatch at a, T_{a1-}, but both of those random variables are used regardless of whether a target is present at location b.

Under reasonable assumptions, this paradigm leads to empirically distinguishable parallel and serial model mean RTs.[5] In particular, if processing is self-terminating, and either the processing time of each probe is equal in distribution across location but not equal in distribution across target/distractor status in the parallel model, or if the mean RTs are unequal across target/distractor status in both positions in the serial model, then serial models make specific predictions that cannot be satisfied by parallel models.

Under these assumptions, the serial model predicts that the sum of the mean RTs across the two trial types in PST-*loc* is equal to the mean RT when both probes are distractors in the PST-*and* added to the mean RT when both probes are targets in the PST-*or*. Under the same assumptions, the parallel model predicts those two sums to be *unequal*. The gist of these inferences can be readily seen using the equations above; for complete proofs, see Townsend and Ashby (1983, Chapter 13).

Using the same paradigm, a slightly different set of assumptions can lead to the same discriminable model predictions as well. One sufficient condition is that the parallel models have unequal first stage target and distractor processing-time distributions for all possible

RTs in both positions; that is, for all $t > 0$,
$$G_{a1+}(t) \neq G_{a1-}(t) \quad \text{and} \quad G_{b1+}(t) \neq G_{b1-}(t).$$
An alternative sufficient condition is that the serial model mean RTs are unequal across targets and distractors in both positions, and $E[T_{b1+}] \neq E[T_{b1-}]$, while the parallel-model distributions are either equal for all possible RTs $t > 0$, and $G_{b1+}^*(t) = G_{b1-}^*(t)$, or unequal for all possible RTs within each position during the first stage, or and $G_{b1+}^*(t) > G_{b1-}^*(t)$ or $G_{b1+}^*(t) < G_{b1-}^*(t)$.

Although the PST has been used relatively much less frequently than the methods described in the next section, it has been successfully applied in a number of domains, including those outside of cognitive psychology. For example, Neufeld and McCarty (1994) used the PST to determine that increased stress levels do not generally lead to higher likelihood of serial processing, regardless of stress susceptibility.

STRONG EXPERIMENTAL TESTS ON RESPONSE TIMES II: APPLICATIONS OF SELECTIVE INFLUENCE

Perhaps the most experimentally popular and also the most technically advanced strategies are those based on what we have termed *systems factorial technology* (Townsend & Nozawa, 1995). Of all the strong tests of hypotheses regarding processing architecture, this one has witnessed a degree of sophisticated investigation and expansion that is almost unprecedented for a young science like psychology. It has also seen wide-ranging applications in a number of diverse spheres of psychological research.

Sternberg, who was responsible for the rebirth and expansion of the Dondersian and Wundt style programs mentioned in the introduction, also invented the *additive factors method*, the direct predecessor to systems factorial technology. Sternberg's striking

---

[5]Note that implicit in this notation is the *selective influence* (defined in the next section) of target presence in a given location on the processing time of that location.

insight was to postulate that for two processes, a and b, there could be two experimental factors, A and B, each of which solely influenced its namesake process. In particular, he supposed that the associated processing time means $E(T_a|A)$ and $E(T_b|B)$ could be sped up or slowed down as functions of A and B, respectively. The reader should observe that $E(T_a|A)$ is completely unaffected by changes in B and vice versa for $E(T_b|B)$. This was the first incarnation of the notion of *selective influence*. With this assumption in place, it was then straightforward to show that a serial system, with a processed before b, predicts the overall mean RT to be $E(T_a + T_b + T_0|A, B] = E(T_a|A) + E(T_b|B) + E(T_0)$. This nice result shows that the mean RT will be an additive function of the experimental factors A, B. The logic was that if an experimenter found additive mean RTs, then the inference was that processing was serial (see Figure 11.2).

Ashby and Townsend (1980) extended the additive factors methods to a distribution-level test. They demonstrated that a serial exhaustive system with factors that are selectively influencing each stage will lead to equal distributions of the sum of RTs when

both levels are the same and distributions of the sum of RTs when levels are mixed. That is, if $RT_{ab;AB}$ is the RT when the factor influencing process a is at level A and the factor influencing process b is at level B, then

$$RT_{ab;11} + RT_{ab;22} = RT_{ab;12} + RT_{ab;21}$$

or, equivalently, with * indicating convolution,

$$g_{ab;11} * g_{ab;22} = g_{ab;12} * g_{ab;21.}$$

This provides an even stronger test for the serial exhaustive model as this equality clearly implies equality at the level of the mean (by the linearity of the expectation operator). Despite the relatively strong implication of the extended additive factors method, this test has indicated the serial exhaustive models in a range of applications. For example, Roberts and Sternberg (1993) demonstrated nearly perfect additivity at the distributional level in a simple detection task and in an identification task.

The tests from Sternberg (1969) and Ashby and Townsend (1980) are powerful. However, if additivity was not supported, then either the architecture was not serial (although what it was, remained unclear),
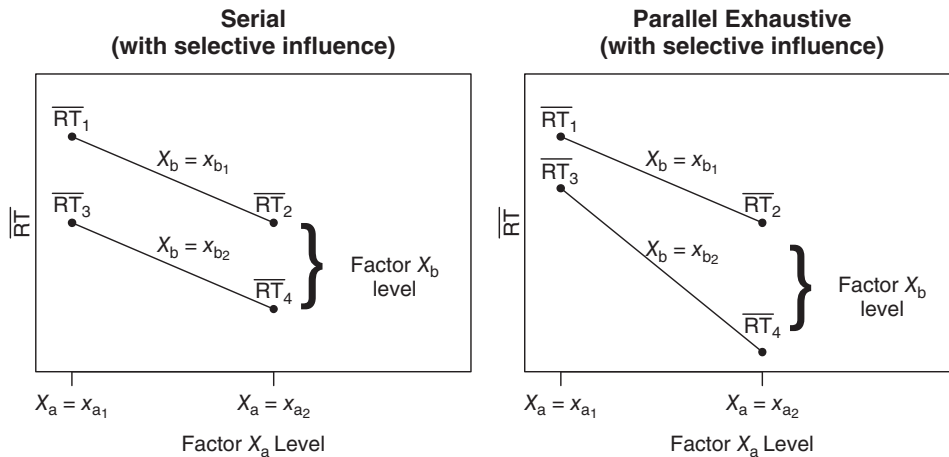

Figure 11.2 Predicted mean response times under selective influence manipulations of each subprocess in a serial system (left) and parallel system (right).

or it might have been serial, but selective influence was not in force. In addition to the limitation of the strict part of the logic to serial architectures, other gaps lay in the absence of a firm technical foundation of the selective influence assumption, and of means of testing its satisfaction. We can only give a somewhat simplified description of the current state of the art, but it should suffice to comprehend our discussions herein.

Most significantly, the theory-driven factorial methodology has been expanded to include (a) tests based on entire RT distributions rather than only means (e.g., Townsend & Nozawa, 1995; Townsend & Wenger, 2004a); (b) serial and parallel architectures along with a broad class of more complex systems (e.g., Chapter 6 in Schweickert, Fisher, & Sung, 2012);[6] and (c) several decisional stopping rules, also identified along with the attendant architecture (e.g., Townsend & Wenger, 2004a).

Definition 7 ensures that (a) effective speeding or slowing of the processes occurs, and (b) the effect is of sufficiently strong form to enable theorems to be proven, which generates powerful tests of the architectures and stopping rules. Even theorems for mean RT interactions currently require selective influence to act at the distribution-ordering level. The reader is referred to Townsend (1984a, 1990) for more detail on this form of selective influence, and Dzhafarov (2003); Townsend, Liu, and Zhang (2017); and Schweickert, Fisher, and Sung (2012) for generalizations and discussions of not only psychological implications but for connections with other central concepts in science in general.

Definition 7: *Distribution-ordering selective influence* holds for processes *a* and *b* with regard to experimental factors *A* and *B*, which themselves form a simple order, if and only if $F_a(t|A_1) = P(T_a \leq t|A_1) < P(T_a \leq t|A_2) = F_a(t|A_2)$ if and only if $A_1 < A_2$ and similarly for $F_b(t|B)$. And, *A* has no effect on $F_b(t|B)$ nor does *B* have any effect on $F_a(t|A)$.

The double factorial paradigm (DFP; Townsend & Nozawa, 1995) offers a further generalization of the additive factors method using distribution-ordering selective influence. The main motivation of the DFP is to isolate the effects of workload on RTs from the effects of architecture and the stopping rule but nonetheless test all three constructs within one paradigm. The first manipulation is to factorially combine the presence or absence of each source of information. Hence, there are trials in which each source is presented in isolation and trials in which all sources are presented together. The second manipulation is meant to factorially speed up and slow down the processing of each present source of information. This manipulation is often referred to as the *salience* manipulation, in reference to the fact that low-salience perceptual stimuli are slower to process than high-salience stimuli, but the manipulations need not necessarily be of salience. Figure 11.3 depicts the possible trial types for an experiment that requires a participant to detect a target either to the right or left of center, or in both locations. Targets are made lower salience on the screen by decreasing the contrast.

Assuming distribution-ordering selective influence of the salience manipulations for each source of information when that source is present, serial and parallel models combined with either exhaustive or first-terminating stopping rules can be discriminated using an interaction contrast of the RT distributions. To maintain consistency

---

[6]A basic assumption of the latter is that the graphs of the connections of the underlying processes are forward flow only. No feedback is generally assumed, although some exceptions can be encompassed (see discussion in Schweickert, Fisher, & Sung, 2012).
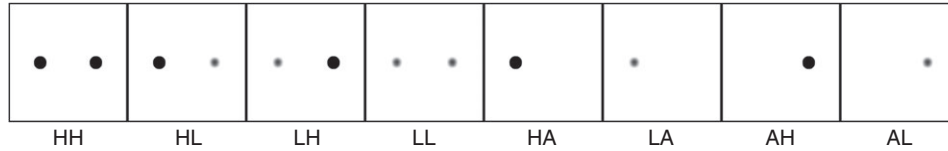
| HH | HL | LH | LL | HA | LA | AH | AL |

Figure 11.3 Example of stimuli for a double factorial experiment to test the architecture and stopping rule of detecting a left and/or right dot. H refers to a high salience level, L refers to a low salience level, and A refers to the absence of the target.

with other papers on systems factorial technology, we switch from the G* to $S(t) = 1 - F(t)$ for the survivor function—that is, the complement of the cumulative distribution function. The survivor interaction contrast is given by

$$SIC(t) = [S_{LL}(t) - S_{HL}(t)] - [S_{LH}(t) - S_{HH}(t)].$$

If distribution-ordering selective influence of the salience manipulation holds, then the survivor functions will be ordered such that when both processes are slowed down, the survivor function will be higher than when either or both of the sources are sped up (i.e., the probability of a RT not occurring yet will always be highest when both sources are slowed down, $(S_{LL}(t) > S_{LH}(t), S_{HL}(t), S_{HH}(t))$. Likewise, the survivor function will be at its lowest when both sources are sped up $(S_{HH}(t) < S_{LH}(t), S_{HL}(t), S_{LL}(t))$.

We have already stated that the salience manipulation, as long as it selectively

influences the processes, is sufficient for discriminating among serial, parallel, exhaustive, and first-terminating models, but why is this the case? Detailed proofs are available in Townsend and Nozawa (1995; for generalizations and alternative approaches see Dzhafarov, Schweickert, & Sung, 2004; Zhang & Dzhafarov, 2015). We build basic intuition for the main results here, and summarize them graphically in Figure 11.4.

First, consider the parallel, first-terminating model. In this model, the fastest process to terminate determines the RT. Hence the overall system will be relatively fast as long as one or both sources are high salience. In terms of the survivor interaction contrast (SIC), this means the first difference, $[S_{LL}(t) - S_{HL}(t)]$, will be relatively larger than the second difference, because only the first difference contains a term without a high salience. Because both terms must be positive (due to the effective selective influence assumption) and the first difference is a large magnitude, we see that the SIC for
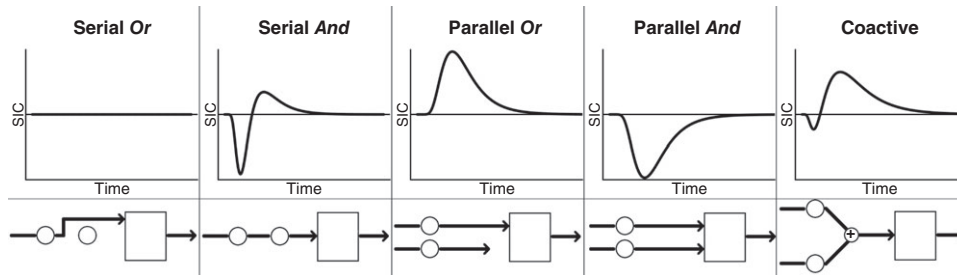

Figure 11.4 Schematic representation of the survivor interaction contrast functions (upper panels) predicted by the models associated with the combination of parallel and serial architectures and *or* and *and* stopping rules, along with predictions for the coactive model.

the parallel, first-terminating model should be positive.

The logic behind the parallel, exhaustive model is similar, but reversed. This model's RT is determined by the last process to finish. Thus the model will be relatively slow if either of the sources are low salience. Hence the second difference in the SIC will be larger than the first difference, because only the second difference contains a term without low salience. Thus, the SIC for the exhaustive model should be negative.

Next, we consider the serial, first-terminating model. In these models, the process to complete first, and hence the one that determines the model completion time, is probabilistically chosen without dependence on the salience levels. The magnitude of the difference between *and* is determined by the probability that the first position is processed first (if the second position is processed first, then the first position is not processed and there is no difference in processing times) and the magnitude of the difference between the low and high salience processing times of the first position. Because both the probability of the first position being processed first and the magnitude of the salience manipulation are the same for both the first and second parts of the SIC, the overall SIC is thus 0 for all times. The serial, exhaustive model is a bit more complicated. The intuition that the overall area under the SIC curve is 0 follows from the fact that the combination of the processing times is additive in this model; hence any manipulation of the individual processing times will result in additivity (i.e., 0 mean interaction). Indeed, this is essentially the result summarized above from Ashby and Townsend (1980). Townsend and Nozawa (1995) go on to show that the SIC is negative for the earliest RTs, and Yang, Fifić, and Townsend (2014) later proved that, as long as at least one of the completion time survivor functions is log concave, there will only be a single 0 crossing for the SIC.[7]

STRONG EXPERIMENTAL TESTS ON RESPONSE FREQUENCIES

Back to the State Spaces: Tests Based on Partial States of Completion

As is the case for the RT literature, there are both strengths and weaknesses to the unique focus on response frequencies; indeed, much of our more recent work has been driven by the potential suggested by using the complementary strengths of each approach (e.g., Eidels, Townsend, Hughes, & Perry, 2015). A branch of theory and methodology pertaining to response frequencies also depends, like the methods in the preceding section, on finer levels of description of the event spaces.

The reader may recall that the PST relied on the fact that parallel systems allow for intermediate stages of processing for any number of uncompleted items, whereas serial systems only permit one item, at most, to be in a partial state of completion. In the present section, we find that this distinction can be utilized in an experimental design on response frequencies, rather than RTs, to test parallel versus serial processing.

The experimental design requires low-to-moderate accuracy, and the observer is instructed to give two responses: his or her first, and a second "guess," just in case his or her first response is incorrect. Parallel models can predict that second guesses are correct at a higher level than chance. Serial models, by virtue of allowing at most one item in a state of partial completion, are severely

---

[7]For the interested reader, Houpt, Blaha, McIntire, Havig, and Townsend (2014) introduced an R package (available on CRAN: https://cran.r-project.org/web/packages/sft/index.html) and a more thorough tutorial for the DFP and the statistical analyses of the associated measures.

restricted on second responses. Following this logic, Townsend and Evans (1983) calculated nonparametric bounds on the number of correct second guesses that any serial model could predict and carried out a preliminary study using this paradigm. Although certain conditions and observers were ruled to have been parallel processing, for a number of observers and conditions, serial processing could not be ruled out. However, only two items were employed, and the serial bounds were not very tight for this small number. This technique should be further exploited, but what has been done so far is in favor of parallel processing (Townsend & Evans, 1983; Van Zandt, 1988). We examine this type of design further next.

Manipulating Process Durations of Available Information

As noted above, consideration of the constellation of issues surrounding the identifiability of serial and parallel systems has not been restricted to RTs. A much smaller, but still substantial, literature on the use of response frequencies (including issues linking process characteristics to signal detection measures of response bias, e.g., Balakrishnan, 1999) has attempted to speak to the issue as well, even above and beyond the above methods based on the differing state spaces of parallel and serial systems. Nominally, the most frequent issue to be addressed by the use of response frequencies has been that of processing capacity and its relation to both process architecture and processing independence. With respect to the latter of these questions, Townsend (1981) considered a potential ambiguity presented by whole-report results of Sperling. Although the serial position data of Sperling seemed to support serial processing with an approximately 10 ms/item scan rate, Townsend noted (as Sperling had acknowledged) that the form

of Sperling's serial position curves suggested the possibility of parallel processing.

Townsend's analysis of the problem presented by these results hinged on the fact that serial processing models based on Poisson distributions all predicted positive item dependencies. As such, a possibility for adjudicating the question of architecture in this context emerged in the potential contrasts between models positing item interdependencies and models assuming independence.

Townsend (1981) developed a set of serial processing models for these comparisons. The first assumed strict serial processing, assuming Poisson distributions for item completion. The second allowed for a random order of serial processing, again, based on Poisson distributions. The third instantiated a capacity limitation in the form of a fixed sample size for processing. This model is an example of a quite popular model for short-term memory called a *slot model*. A fourth assumed independent parallel channels. The first two of these models were shown to predict positive item interdependencies. The third, a general slot model, predicted negative dependencies. The logic is straightforward: Slot models assume there exist a finite number of slots available for item storage. The more items already occupying a slot, the less chance another item has to find a free slot. Even assuming that the number of slots can vary from trial to trial does not help—the result is simply a mixture of slot sizes, each of which predicts negative dependence, as does the overall mixture. The most natural architecture for an independence model is parallel, as noted above, although other architectures are possible (see, e.g., Townsend & Ashby, 1983, Chapter 4).

When the data from Townsend's (1981) whole-report study were plotted against the predictions of these competing models, it revealed a high level of agreement between

the data and the predictions of the model that assumed independence, consistent with the idea of parallel processing. Further, Townsend (1981) considered two possible forms of parallel processing—one assuming independent rates of information accumulation across positions in the display and a second assuming limitations in the allocation of processing resources—and showed that the former offered the superior fit to the data.

Subsequently, possibly the most frequently used paradigm to explore questions regarding processing characteristics using response frequencies has been some variant on what is referred to as the simultaneous-sequential paradigm. This paradigm, introduced in its original form by Eriksen and Spencer (1969), has been widely used to examine questions regarding processing capacity at the individual item level (e.g., Duncan, 1980; Fisher, 1984; Harris, Pashler, & Coburn, 2004; Hoffman, 1978; Huang & Pashler, 2005; Kleiss & Lane, 1986; Prinzmetal & Banks, 1983; Shiffrin & Gardner, 1972). In its typical form, the simultaneous-sequential paradigm involves the use of varying numbers of stimuli in two conditions, one in which all of the stimuli are presented at the same time (simultaneous condition) and one in which they are presented one at a time (sequential). The frame duration, which can be manipulated across a range of values, is held constant across the two display conditions.[8]

In the most felicitous paradigm of this nature, the experimenter acquires some preliminary data in order to estimate the average duration consumed by the processing (e.g., search, identification, etc.) of a single item. Suppose that time is found to

be approximately $t*$ ms. Thus, we have the basic setup that one condition presents all n items simultaneously for $t*$ ms, whereas the other condition presents each item for $t*$ ms, each in succession. Obviously, a standard parallel system should not be seriously hurt in performance in the simultaneous condition as opposed to the sequential condition. In contrast, a standard serial system should only be able to process approximately one item in that condition versus about n in the sequential condition.

Eriksen and Spencer's (1969) introduction of the paradigm was motivated by a very careful review of the range of issues that had to be considered in order for an experiment to convincingly address the question of how much information can be extracted from a stimulus display in a unit of time (capacity), along with how the information is extracted (serially or in parallel; see Estes & Taylor, 1964, 1966; Estes & Wessel, 1966; Sperling, 1963). The potentially complicating factors considered by Eriksen and Spencer (1969) included capacity limitations at the levels of low-level stimulus features (duration, noise, boundaries), low-level perceptual characteristics (foveal), and postperceptual, higher level aspects of attention and memory.

As Eriksen and Spencer (1969, p. 2) noted, echoing the issues we have discussed above in the context of RTs, "Definitive answers [regarding capacity and process architecture] have not been obtained primarily because of the numerous problems that are involved in what at first glance seems to be a simple methodology. The difficulties that are encountered become apparent if the input-output information flow is considered in terms of the various distinguishable subprocesses that are involved." Yet as they noted, there seemed to be strong and accumulating evidence (e.g., Eriksen & Lappin, 1965) that there must be some capacity limitations at some level of processing, with

[8]In the actual experiment, Eriksen and Spencer (1969) only approximated the simultaneous condition by extremely rapid sequential presentation of the items. However, this manipulation apparently was perceptually equivalent to simultaneous presentation.

the general approach at the time suggesting a mapping between capacity limitations and serial processing.

Eriksen and Spencer's (1969) results, in this context, were quite simple and compelling. Essentially, the only experimental factor among the range of those manipulated to have any regular effect on observer's accuracy (quantified in terms of signal detection theory) was the number of items in the display set, with there being a monotonic reduction in sensitivity from one item ($d'_A = 1.64$) to nine items ($d'_A = 0.88$), with this loss of sensitivity being due almost entirely to an increase in the rate of false positives. Eriksen and Spencer's explanation for these results was parsimonious and intriguingly similar to Sternberg's logic. Specifically, they reasoned that if capacity was constant as a function of display size, the hit and false positive rates for single-item displays could be used, under the assumption of independence of item processing, to predict the level of performance across variations in display size. Their predictions were highly consistent with their data. Their conclusion illustrates both the potency of this very simple and elegant reasoning on probabilities and the contemporary conflation of the issues of capacity and processing architecture (p. 15; for related research, see Averbach & Coriell, 1961; Mayzner, Tresselt, & Helper, 1967; Sperling, 1963): "This would suggest that the encoding mechanism proposed in current theories…can scan through or encode nine letters as efficiently in 50 msec. as in 25 msec. This interpretation would almost certainly preclude a serial encoding process."

Townsend and colleagues (Townsend, 1981; Townsend & Ashby, 1983, Chapter 11; Townsend & Fial, 1968) extended the Eriksen and Spencer logic in a visual whole-report setting illustrated in Figure 11.5. Visual whole reports ask the observer to report all
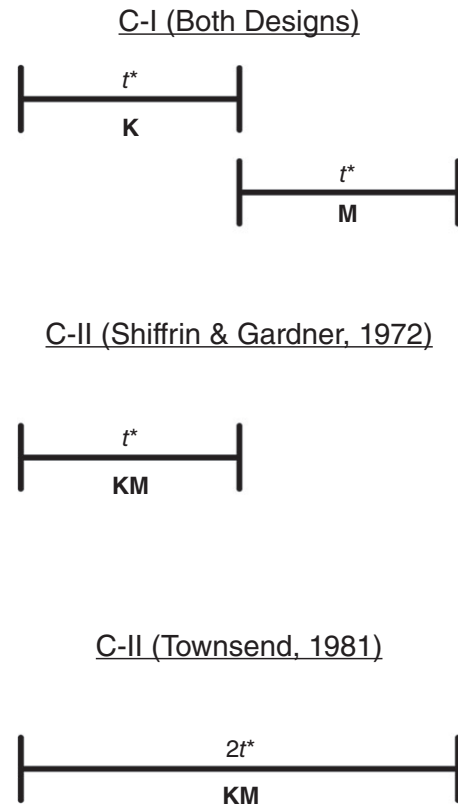

Figure 11.5   Schematic of two related designs for discriminating parallel and serial processing based on accuracy. In both designs, participants are shown a fixed number of items, then asked to report as many as possible. A baseline of accuracy is estimated by presenting each item in isolation for t^* ms. In the Shiffrin and Gardner (1972) design, the baseline accuracy is compared against performance when all items are presented together for t^* ms. In the Townsend (1981) design, all items are presented together for n times t^* ms (where n is the number of items).

the items she or he is able to from a visual display. Instead of the t^* ms simultaneous condition versus $n \times t^*$ successive item intervals, as in the original Eriksen and Spencer (1969), or that of Shiffrin and Gardner's (1972) design (see below), Townsend and colleagues added a pair of conditions, with one condition being the same; that is, n items being presented sequentially for t^*

ms each. This was compared with a new condition where all *n* items were presented simultaneously for $n \times t^*$ ms. In contrast to the original design, here a standard serial model is predicted to do as well, but no better, in the simultaneous versus the sequential condition. On the other hand, the parallel system, assuming it's not working at ceiling level, is expected to perform much better in the simultaneous condition. The results, like those of their independence analyses, strongly supported parallel processing channels.

Shiffrin and Gardner (1972) employed the original Eriksen and Spencer (1969) idea. They found evidence of equality of performance in the two presentation conditions across a set of variations, leading them to a conclusion consistent with that of Eriksen and Spencer: specifically, that processing was parallel and unlimited in capacity (and by extension, parallel) up to the level of postperceptual processing.

More recently, Palmer and colleagues (e.g., Scharff, Palmer, & Moore, 2011) have extended the simultaneous-sequential paradigm to broaden the class of models of capacity that can be addressed in a single experiment. Specifically, they added a condition in which all of the test stimuli are in two frames with equal durations, separated by a 1,000-ms blank interval intended to allow sufficient time for attention switching (Ward, Duncan, & Shapiro, 1996). Recall that a switching interval "recess" was not present in any of the earlier applications of this strategy. They then used a set of probability models to derive contrasting predictions for the various experimental conditions in their extended paradigm.

Most critically, their predictions were even stronger than those associated with previous uses of the paradigm. The predictions contrast three classes of models (Scharff et al., 2011, p. 816) across specific subconditions of their expanded simultaneous-sequential paradigm. The three classes of models are referred to as limited, intermediate, and unlimited capacity models. We refer readers to Scharff et al.'s (2011) paper for the specific predictions, noting here that the logic supporting those predictions does allow, under specific assumptions regarding processing architecture and independence for both confirmation and falsification of competing accounts, underscoring the recurring theme that these issues cannot be considered in isolation.

CONCLUSION

As we asserted earlier, it is our belief that almost all of psychological science can, and should be, thought of as a black box discipline. Such a discipline would be analogous to the several engineering, computer science, and applied physics fields that attempt, largely through analysis of behavioral input-output regularities, to discover, affirm, or disconfirm internal systems that can produce the regularities. We can likely borrow at least some elements directly from these fields, such as the theory of the ideal detector from mathematical communications science. However, because of the complexity of the brain and the perversely constrained means we have of studying even input-output relationships, the greater part of such strategies must be vastly different in detail from those of these other sciences.

We hasten to emphasize that this type of program is hardly *behavioristic* in the traditional Skinnerian sense of the term.[9]

---

[9] An intriguing interchange with B. F. Skinner in his later days and a wide range of scientists can be found in B. F. Skinner's "Methods and Theories in the Experimental Analysis of Behavior" (1984). In particular, it contains a small debate between Skinner and Townsend on the possibilities of mathematical modeling of human psychology (Townsend, 1984b).

Rather than logging the entries into a gargantuan dictionary of stimulus-response correlations, as in Skinner's early agenda, the overriding goal of our type of program is to move from the input-output regularities, garnered from specially designed experimental paradigms, to sets of internal structures and processes.[10] Most, if not all, of our approaches attempt to thoroughly adhere to the concept of a psychological systems science, and, in the bargain, to make ourselves aware of when model mimicking is likely to occur and how to avoid it by metamodeling deeper experiment methods of assessing important processing mechanisms. The potency of this adherence to a systems view allows for application to a range of psychological phenomena, not just those in the areas of perception, memory, and perceptual organization in which we have worked. Space prevents us from giving even a brief description to this range of studies; however, we supply in the appendix a listing of applications of aspects of our approach.

The preponderance of methods introduced in this chapter use RTs as the major observable variable, although we saw several based on response frequencies, and therefore inclusive of accuracy. A major facet of our armamentarium that we have had to neglect in the present treatment has employed RTs to measure workload capacity, the way in which efficiency changes as the workload n is increased (see, e.g., Townsend &

Nozawa, 1995; Townsend & Wenger, 2004a, 2004b). For instance, the change in RTs with increases in memory-set size in our prototypical Sternberg (1966) paradigm is a good example. Our statistic $C(t)$ provides a measure where performance as a function of n is compared against a baseline predicted by unlimited capacity independent parallel processing. Very recently, we have generalized the $C(t)$ statistic to take into account accuracy as well as RTs in measuring performance as a function of n. The new statistic is called $A(t)$, for *assessment function* (Townsend & Altieri, 2012).

A "must do" extension of our methodologies is to systems of arbitrary number n of operational processes, rather than just $n = 2$. This target has lately been met for serial versus parallel systems and for exhaustive and minimum time-stopping rules (Yang, Fifić, & Townsend, 2014). Another highly valuable theme of our methodology, one that was originally based on accuracy and confusion frequencies alone and was co-invented with F. G. Ashby, is *general recognition theory* (Ashby & Townsend, 1986). Over the years Ashby and colleagues have developed special models that bring together concepts from general recognition theory and RT processing (Ashby, 1989, 2000). They also extended general recognition theory into a powerful theory of categorization (e.g., Ashby, 1992).

We have recently developed a very general class of models entitled *response time general recognition theory* (Townsend, Houpt, & Silbert, 2012). Currently, these models can employ both accurate responses and various types of confusions in concert with the associated RTs in order to assess perceptual and cognitive dependencies, as well as higher order classification, at the deepest conceivable level. However, at present our methodology does not permit tests of architecture or stopping rules, as does systems factorial technology, with RTs as the

---

[10]In the advent of early cognitive science in the 1960s, it was fashionable to vilify behaviorism as a dark side of scientific enterprise. In truth, it can be forcefully argued that a few decades of rather strict behaviorism, along with the logical positivism of the Vienna Circle and friends, and the allied new emphasis, even in physics, of operationalism, was important in helping psychology to finish pruning itself from its mother philosophical roots. Furthermore, in careful reading of the works of the great theoretical behaviorist of the day, Clark Hull, the rudiments of modern computational and mathematical psychology are readily visible.

dependent variable. Closer to our exposition here, we are currently extending systems factorial technology to incorporate response frequencies in identification of architecture and stopping rules.

There is an important omission from our current set of topics. Virtually all of the systems under study here obey what has been called the *Dondersian postulate* (Townsend & Ashby, 1983, pp. 5–6, 358, 408); namely, that each psychological process in a sequence of such processes finishes before the next in the chain begins. A traditional supplement to the discrete flow assumption is the assumption of pure insertion, mentioned earlier in this chapter, wherein the experimenter can insert or withdraw a stage or process by some type of empirical manipulation. However, this additional restriction is not mandatory. A more suggestive name for the Dondersian assumption is *discrete flow*.

However, anyone who has studied natural systems (for example), as approached or described with differential equations, is aware that many physical, chemical, and biological systems obey the opposite precept of *continuous flow*. The output of a process in a continuous flow system is instantaneously input to the succeeding process. And, the input space, state space, and output space are usually depicted as continuous functions (e.g., see Townsend & Ashby, 1983, pp. 401–419). These systems are sometimes called *lumped systems* in engineering because there is no time lag at all from one stage to the next.

The simplest such systems, still of extreme importance in the sciences, are linear in nature. Even the output of linear systems, however, can reflect memory of past inputs. A relatively early example of a continuous flow linear system was the cascade model introduced by McClelland (1979). Within systems factorial technology, even though this type of model does not predict additivity

in mean RTs as do true serial models that obey selective influence, McClelland showed that it can approximate such additivity.

The amount of theoretical knowledge concerning, for instance, how to identify such spaces and when one might predict additivity of factorial effects in an observable state space is very modest indeed. Schweickert and colleagues (1989; Schweickert & Mounts, 1998) have derived results for certain continuous flow systems that include, but are not limited to, linear systems. However, their targeted systems do assume a lack of memory on past inputs or states. Townsend and Fikes (1995) studied a broad class of possibly nonlinear, continuous flow systems that do include such memories across time. Miller (1988) has investigated hybrid systems that may only partially process items and possess temporal overlap analogous to continuous flow systems.

The deeper analysis of continuous flow and hybrid systems and development of tools for their identification should be considered a top priority for psychological theorists.

At some level, all of the theory and application that we have considered in this chapter must make some form of contact with measurable aspects of the activity of the nervous system. Although quite limited at the moment, there are examples of how the foundational ideas considered here might be related to neurophysiology. Possibly the earliest consideration of these possibilities can be found in Schweickert and colleagues' work on continuous flow systems as mentioned above (1989; Schweickert & Mounts, 1998), where consideration was given to the use of factorial methods in application to event-related potentials.

Since that time, a handful of studies have considered the utility of the additive factors method with respect to both event-related potentials (e.g., Miller & Hackley, 1992) and the blood-oxygen level dependent signal

in magnetic resonance imaging (e.g., Cummine, Sarty, & Borokowski, 2010), and the utility of the capacity measures in conjunction with an electroencephalogram (Blaha, Busey, & Townsend, 2009). To our knowledge, however, there is at present only one as-yet unpublished study (initial reports in Wenger, Ingvalson, & Rhoten, 2017; Wenger & Rhoten, 2015) that has attempted to apply the complete set of measures associated with systems factorial technology to neurophysiological data. In that study, the onset time of the lateralized readiness potential (an electroencephalogram feature that reliably precedes choice RT; see, for example, Mordkoff & Grosjean, 2001) was used as the dependent variable. Conclusions drawn from the analyses of these data were consistent, on a subject-by-subject basis, with the conclusions drawn from the RT data.

Although preliminary, these results are quite encouraging with respect to one possible way of connecting extant theory and method with cognitive neuroscience.

The ultimate goal of all of these endeavors is to have in hand a powerful armamentarium whose tools can utilize both of the strongest observable dependent variables available: response frequencies and RTs. These tools are expected to simultaneously assess architecture, stopping rules, attentional and other kinds of capacity, and finally, several types of process and item dependencies. In addition, it is our hope that we can join with neuroscientists in employing our methodologies along with their diverse strategies, to pinpoint underlying mechanisms and their interactions. As we peer into the future, we envision researchers from many cognate fields developing rigorous varieties of psychological systems theory, ones that are general (i.e., not glued to specific aspects of processing, such as particular probability distributions) and conversant with the challenges of model mimicking.

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. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford handbook of computational and mathematical psychology* (pp. 63–98). New York, NY: Oxford University Press.

Ashby, F. G. (1989). Stochastic general recognition theory. In D. Vickers & P. L. Smith (Eds.), *Human information processing: Measures, mechanisms, and models: Proceedings of the 24th International Congress of Psychology of the International Union of Psychological Science* (Vol. 2, pp. 435–457). Sydney, Australia: International Union of Psychological Science.

Ashby, F. G. (1992). Multidimensional models of categorization. In S. W. Link & J. T. Townsend (Series Eds.) & F. G. Ashby (Vol. Ed.), *Scientific psychology series: Multidimensional models of perception and cognition* (pp. 449–483). Hillsdale, NJ: Erlbaum.

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, *44*(2), 310–329.

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

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154.

Atkinson, R. C., Holmgren, J. E., & Juola, J. F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics*, *6*, 321–326.

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal*, *40*(1), 309–328.

Balakrishnan, J. D. (1999). Decision processes in discrimination: Fundamental misrepresentations of signal detection theory. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(5), 1189.

Bamber, D. (1969). Reaction times and error rates for "same"-"different" judgments of multidimensional stimuli. *Perception & Psychophysics*, *6*, 169–174.

Billingsley, P. (1986). *Probability and measure*. New York, NY: Wiley.

Blaha, L. M., Busey, T. A., & Townsend, J. T. (2009). An LDA approach to the neural correlates of configural learning. In N. Taatgen (Ed.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Booth, T. L. (1967). *Sequential machines and automata theory*. New York, NY: Wiley.

Busemeyer, J. B., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.

Chung, K. L. (2001). *A course in probability theory*. San Diego, CA: Academic Press.

Cummine, J., Sarty, G. E., & Borowsky, R. (2010). Localizing the Frequency × Regularity word reading interaction in the cerebral cortex. *Neuropsychologia*, *48*(7), 2147–2157.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*, 272–300.

Dzhafarov, E. N. (1993). Grice-representability of response time distribution families. *Psychometrika*, *58*(2), 281–314.

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*, 245–252.

Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychonomic Bulletin & Review*, *17*(6), 763–771.

Eidels, A., Townsend, J. T., Hughes, H. C., & Perry, L. A. (2015). Evaluating perceptual integration: Uniting response-time- and accuracy-based methodologies. *Attention, Perception, & Psychophysics*, *77*, 659–680.

Eriksen, C. W., & Lappin, J. S. (1965). Internal perceptual system noise and redundancy in simultaneous inputs in form identification. *Psychonomic Science*, *2*, 351–352.

Eriksen, C. W., & Spencer, T. (1969). Rate of information processing in visual perception: Some results and methodological considerations. *Journal of Experimental Psychology*, *79*(2), 1–16.

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, USA*, *52*(2), 446–454.

Estes, W. K., & Taylor, H. A. (1966). Visual detection in relation to display size and redundancy of critical elements I. *Attention, Perception, & Psychophysics*, *1*(1), 9–16.

Estes, W. K., & Wessel, D. L. (1966). Reaction time in relation to display size and correctness of response in forced-choice visual signal detection. *Perception & Psychophysics*, *1*(11), 369–373.

Fisher, S. (1984). Central capacity limits in consistent mapping, visual search tasks: Four channels or more? *Cognitive Psychology*, *16*, 449–484.

Greeno, J. G., & Steiner, T. E. (1964). Markovian processes with identifiable states: General considerations and application to all-or-none learning. *Psychometrika*, *29*(4), 309–333.

Greenwald, A. G. (2012). There is nothing so theoretical as a good method. *Perspectives on Psychological Science*, *7*, 99–108.

Harris, C. R., Pashler, H. E., & Coburn, N. (2004). Moray revisited: High-priority affective stimuli and visual search. *Quarterly Journal of Experimental Psychology*, *57A*(1), 1–31.

Hoffman, J. E. (1978). Search through a sequentially presented visual display. *Perception & Psychophysics*, *23*, 1–11.

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., & Townsend, J. T. (2014). Systems

factorial technology with R. *Behavior Research Methods*, *46*, 307–330.

Houpt, J. W., Townsend, J. T., & Jefferson, B. (2017). Stochastic foundations of elementary mental architectures. In *New Handbook of Mathematical Psychology*, *Vol. 2*. Cambridge University Press.

Huang, L., & Pashler, H. (2005). Attention capacity and task difficulty in visual search. *Cognition*, *94*, B101–B111.

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, *121*(1), 1–32.

Khodadadi, A., & Townsend, J. T. (2015). On mimicry among sequential sampling models. *Journal of Mathematical Psychology, 68–69*, 37–48.

Kleiss, J. A., & Lane, D. M. (1986). Locus and persistence of capacity limitations in visual information processing. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 200–210.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York, NY: Academic Press.

Krueger, L. E. (1978). A theory of perceptual matching. *Psychological Review*, *85*(4), 278.

Kolmogorov, A. N. (1950). *Foundations of the Theory of Probability*. New York: Chelsea Press.

Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction*. Hillsdale, NJ: Erlbaum.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, *40*, 77–105.

Mayzner, M. S., Tresselt, M. E., & Helper, M. S. (1967). A provisional model of visual information processing with sequential inputs. *Psychonomic Monograph Supplement*, *2*, 91–108.

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

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

Miller, J., & Hackley, S. A. (1992). Electrophysiological evidence for temporal overlap among contingent mental procesess. *Journal of Experimental Psychology: General*, *121*, 195–209.

Mordkoff, J. T., & Grosjean, M. (2001). The lateralized readiness potential and response kinetics in response-time tasks. *Psychophysiology*, *38*(5), 777–786.

Neufeld, R. W. J., & McCarty, T. S. (1994). A formal anlaysis of stressor and stress-proneness effects on simple information processing. *British Journal of Mathematical and Statistical Psychology*, *47*, 193–226.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing: Proceedings of the Eighth Annual Carnegie Symposium on Cognition* (pp. 283–305). New York, NY: Academic Press.

Nosofsky, R. M., Little, D. R., Donkin, C., & Fifić, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*(2), 280.

Platt, J. R. (1964). Strong inference. *Science*, *146*, 347–353.

Prinzmetal, W., & Banks, W. P. (1983). Perceptual capacity limits in visual detection and search. *Bulletin of the Psychonomic Society*, *4*, 263–266.

Rao, B. L. S. P. (1992). *Identifiability in stochastic models: Characterization of probability distributions*. New York, NY: Academic Press.

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

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 611–654). Cambridge, MA: MIT Press.

Scharff, A., Palmer, J., & Moore, C. M. (2011). Extending the simultaneous-sequential paradigm to measure perceptual capacity for features

and words. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 813.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.

Schweickert, R. (1989). Separable effects of factors on activation functions in discrete and continuous models: d′ and evoked potentials. *Psychological Bulletin*, *106*(2), 318–328.

Schweickert, R., Fisher, D. L., & Sung, K. (2012). *Discovering cognitive architecture by selectively influencing mental processes* (Vol. 4). Hackensack, NJ: World Scientific.

Schweickert, R., & Mounts, J. (1998). Additive effects of factors on reaction time and evoked potentials in continuous flow models. In S. W. Link & J. T. Townsend (Series Eds.) & C. Dowling, F. Roberts, & P. Theuns (Vol. Eds.), *Scientific psychology series: Recent progress in mathematical psychology: Psychophysics, knowledge, representation, cognition and measurement*. Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology*, *93*, 78–82.

Skinner, B. F. (1984). Methods and theories in the experimental analysis of behavior. *Behavioral and Brain Sciences,* *7*(4), 511–523.

Smith, P. L., & Van Zandt, T. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology*, *53*(2), 293–315.

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

Sperling, G. A. (1963). A model for visual memory tasks. *Human Factors*, *5*, 19–31.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652–654.

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

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, *27*(1), 1–32.

Townsend, J. T. (1969, April). *Mock parallel and serial models and experimental detection of these.* Paper presented at the Purdue Centennial Symposium on Information Processing, West Lafayette, IN.

Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, *10*, 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.

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. (1981). Some characteristics of visual whole report behavior. *Acta Psychologica*, *47*, 149–173.

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

Townsend, J. T. (1984b). Psychology: Toward the mathematical inner man. Commentary on Target article by B. F. Skinner: Methods and theories in the experimental analysis of behavior. *Behavioral and Brain Sciences*, *7*, 539–540.

Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, *108*, 551–567.

Townsend, J. T. (1992). On the proper scales for reaction time. In S. W. Link & J. T. Townsend (Series Eds.) & H.-G. Geissler, S. W. Link, &

J. T. Townsend (Vol. Eds.), *Scientific psychology series: Cognition, information processing, and psychophysics: Basic issues* (pp. 105–120). Hillsdale, NJ: Erlbaum.

Townsend, J. T., & Altieri, N. (2012). An accuracy–response time capacity assessment function that measures performance against standard parallel predictions. *Psychological Review*, *119*(3), 500.

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

Townsend, J. T., & Ashby, F. G. (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin*, *90*, 394–401.

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, Germany: VEB Deutscher Verlag der Wissenschaften.

Townsend, J. T., & Fial, R. (1968, August). *Spatio-temporal characteristics of multisymbol perception*. Paper presented at the First Annual Mathematical Psychological Meeting, Stanford, CA.

Townsend, J. T., & Fifić, M. (2004). Parallel versus serial processing and individual differences in high-speed search in human memory. *Perception & Psychophysics*, *66*(6), 953–962.

Townsend, J. T., & Fikes, T. (1995). *A beginning quantitative taxonomy of cognitive activation systems and application to continuous flow processes*. Cognitive Science Research Report 131. Bloomington: Indiana University.

Townsend, J. T., Houpt, J. W., & Silbert, N. H. (2012). General recognition theory extended to include response times: Predictions for a class of parallel systems. *Journal of Mathematical Psychology*, *56*(6), 476–494.

Townsend, J. T., Liu, Y., & Zhang, R. (2017). Selective influence and classificatory separability (perceptual separability) in perception and cognition: Similarities, distinctions, and synthesis. In D. R. Little, N. Altieri, M. Fifić, &

C.-T. Yang (Eds.), *Systems factorology: A theory-driven methodology for the identification of perceptual and cognitive mechanisms* (pp. 93–113). San Diego, CA: Academic Press.

Townsend, J. T., & Nozawa, G. (1995). On the spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*, 321–359.

Townsend, J. T., & Roos, R. N. (1973). Search reaction time for single targets in multiletter stimuli with brief visual displays. *Memory & Cognition*, *1*(3), 319–332.

Townsend, J. T., & Wenger, M. J. (2004a). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review*, *11*(3), 391–418.

Townsend, J. T., & Wenger, M. J. (2004b). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003–1035.

Van Zandt, T. (1988). *Testing serial and parallel processing hypotheses in visual whole report experiments* (Master's thesis). Purdue University, West Lafayette, IN.

Vorberg, D. (1977, August). *On the equivalence of parallel and serial models of information processing*. Paper presented at the 10th Annual Meeting of the Society for Mathematical Psychology, Los Angeles, CA.

Ward, R., Duncan, J., & Shapiro, K. (1996). The slow time-course of visual attention. *Cognitive Psychology*, *30*(1), 79–109.

Wenger, M. J., Ingvalson, E. M., & Rhoten, S. E. (2017). The continuing evolution of systems factorial theory: Connecting theory with behavioral and neural data. In D. R. Little, N. Altieri, M. Fifić, & C.-T. Yang (Eds.), *Systems factorial technology: A theory driven methodology for the identification of perceptual and cognitive mechanisms*. San Diego, CA: Academic Press.

Wenger, M. J., & Rhoten, S. E. (2015, October). *Perceptual learning for detection of multiple features produces non-independent processing: Behavioral and neurophysiological evidence*.

Poster session presented at the 2015 meeting of the Society for Neuroscience, Chicago, IL.

Woodworth, R. S. (1938). *Experimental psychology*. New York, NY: Henry Holt.

Yang, C.-T., Little, D. R., & Hsu, C.-C. (2014). The influence of cueing on attentional focus in perceptual decision making. *Attention, Perception, & Psychophysics*, *76*(8), 2256–2275.

Yang, H., Fifić, M., & Townsend, J. T. (2014). Survivor interaction contrast wiggle 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*, *6*(735).

APPENDIX: APPLICATIONS OF SYSTEMS FACTORIAL TECHNOLOGY

Age-Related Changes in Perception and Cognition

Ben-David, B. M., Eidels, A., & Donkin, C. (2014). Effects of aging and distractors on detection of redundant visual targets and capacity: Do older adults integrate visual targets differently than younger adults? *PLOS ONE*, *9*(12), e113551.

Gottlob, L. R. (2007). Aging and capacity in the same-different judgment. *Aging Neuropsychology and Cognition*, *14*, 55–69.

Wenger, M. J., Negash, S., Petersen, R. C., & Petersen, L. (2010). Modeling and estimating recall processing capacity: Sensitivity and diagnostic utility in application to mild cognitive impairment. *Journal of Mathematical Psychology*, *54*, 73–89.

Binocular Interaction

Hughes, H. C., & Townsend, J. T. (1998). Varieties of binocular interaction in human vision. *Psychological Science*, *9*, 53–60.

Medina, J. M. (2006). Binocular interactions in random chromatic changes at isoluminance. *Journal of the Optical Society of America A: Optics Image Science and Vision*, *23*, 239–246.

Medina, J. M., & Diaz, J. A. (2006). Postreceptoral chromatic-adaptation mechanisms in the red-green and blue-yellow systems using simple reaction times. *Journal of the Optical Society of America A: Optics Image Science and Vision*, *23*, 993–1007.

Categorization

Blunden, A. G., Wang, T., Griffiths, D. W., & Little, D. R. (2015). Logical-rules and the classification of integral dimensions: Individual differences in the processing of arbitrary dimensions. *Frontiers in Psychology*, *5*, 1531.

Fifić, M., Nosofsky, R. M., & Townsend, J. T. (2008). Information-processing architectures in multidimensional classification: A validation test of the systems factorial technology. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 356–375.

Fifić, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-Rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review*, *117*, 309–348.

Little, D. R., Nosofsky, R. M., & Denton, S. E. (2011). Response-Time tests of logical-rule models of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1–27.

Cognitive Control

Eidels, A., Townsend, J. T., & Algom, D. (2010). Comparing perception of Stroop stimuli in focused versus divided attention paradigms: Evidence for dramatic processing differences. *Cognition*, *114*, 129–150.

Human–Machine Teaming

Yamani, Y., & McCarley, J. S. (2016). Workload capacity: A response time-based measure of automation dependence. *Human Factors 58*(3), 462–471.

Identity and Emotion Perception

Fitousi, D., & Wenger, M. J. (2013). Variants of independence in the perception of facial identity and expression. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 133–155.

Richards, H. J., Hadwin, J. A., Benson, V., Wenger, M. J., & Donnelly, N. (2011). The influence of anxiety on processing capacity for threat detection. *Psychonomic Bulletin & Review*, *18*, 883–889.

Yankouskaya, A., Humphreys, G. W., & Rotshtein, P. (2014). Differential interactions between identity and emotional expression in own and other-race faces: Effects of familiarity revealed through redundancy gains. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1025–1038.

Yankouskaya, A., Humphreys, G. W., & Rotshtein, P. (2014). The processing of facial identity and expression is interactive, but dependent on task and experience. *Frontiers in Human Neuroscience*, *8*, 920.

Individual Differences/Clinical Populations

Endres, M. J., Houpt, J. W., Donkin, C., & Finn, P. R. (2015). Working memory capacity and redundant information processing efficiency. *Frontiers in Psychology*, *6*, 594.

Houpt, J. W., Sussman, B. L., Townsend, J. T., & Newman, S. D. (2015). Dyslexia and configural perception of character sequences. *Frontiers in Psychology*, *6*, 482.

Johnson, S. A., Blaha, L. M., Houpt, J. W., & Townsend, J. T. (2010). Systems factorial technology provides new insights on global-local information processing in autism spectrum disorders. *Journal of Mathematical Psychology*, *54*, 53–72.

Neufeld, R. W. J., Vollick, D., Carter, J. R., Boksman, K., & Jette, J. (2002). Application of stochastic modeling to the assessment of group and individual differences in cognitive functioning. *Psychological Assessment*, *14*, 279–298.

Learning and Reward Processing

Sui, J., Yankouskaya, A., & Humphreys, G. W. (2015). Super-Capacity me! Super-Capacity and violations of race independence for self- but not for reward-associated stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 441–452.

Wenger, M. J. (1999). On the whats and hows of retrieval in the acquisition of a simple skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1137–1160.

Memory Search

Donkin, C., & Nosofsky, R. M. (2012). The structure of short-term memory scanning: An investigation using response time distribution models. *Psychonomic Bulletin & Review*, *19*, 363–394.

Thomas, R. (2006). Processing time predictions of current models of perception in the classic additive factors paradigm. *Journal of Mathematical Psychology*, *50*, 441–455.

Townsend, J. T., & Fifić, M. (2004). Parallel versus serial processing and individual differences in high-speed search in human memory. *Perception & Psychophysics*, *66*, 953–962.

Multimodal Interaction

Altieri, N., & Hudock, D. (2014). Hearing impairment and audiovisual speech integration ability: A case study report. *Frontiers in Psychology*, *5*, 678.

Altieri, N., & Hudock, D. (2014). Assessing variability in audiovisual speech integration skills using capacity and accuracy measures. *International Journal of Audiology*, *53*, 710–718.

Altieri, N., Pisoni, D. B., & Townsend, J. T. (2011). Some behavioral and neurobiological constraints on theories of audiovisual speech integration: A review and suggestions for new directions. *Seeing and Perceiving*, *24*, 513–539.

Altieri, N., Stevenson, R. A., Wallace, M. T., & Wenger, M. J. (2015). Learning to associate auditory and visual stimuli: Behavioral and neural mechanisms. *Brain Topography*, *28*, 479–493.

Altieri, N., & Townsend, J. T. (2011). An assessment of behavioral dynamic information processing measures in audiovisual speech perception. *Frontiers in Psychology*, *2*, 238.

Altieri, N., Townsend, J. T., & Wenger, M. J. (2014). A measure for assessing the effects of audiovisual speech integration. *Behavior Research Methods*, *46*, 406–415.

Altieri, N., & Wenger, M. J. (2013). Neural dynamics of audiovisual speech integration under variable listening conditions: An individual participant analysis. *Frontiers in Psychology*, *4*, 615.

Berryhill, M., Kveraga, K., Webb, L., & Hughes, H. C. (2007). Multimodal access to verbal name codes. *Perception & Psychophysics*, *69*, 628–640.

Blurton, S. P., Greenlee, M. W., & Gondan, M. (2014). Multisensory processing of redundant information in go/no-go and choice responses. *Attention, Perception, & Psychophysics, 76*, 1212–1233.

Hochenberger, R., Busch, N. A., & Ohla, K. (2015). Nonlinear response speedup in bimodal visual-olfactory object identification. *Frontiers in Psychology*, *6*, 1477.

Hugenschmidt, C. E., Hayasaka, S., Peiffer, A. M., & Laurienti, P. J. (2010). Applying capacity analyses to psychophysical evaluation of multisensory interactions. *Information Fusion*, *11*, 12–20.

Patching, G. R., & Quinlan, P. T. (2004). Cross-modal integration of simple auditory and visual events. *Perception & Psychophysics*, *66*, 131–140.

Perceptual Organization

Donnelly, N., Cornes, K., & Menneer, T. (2012). An examination of the processing capacity of features in the Thatcher illusion. *Attention, Perception, & Psychophysics*, *74*, 1475–1487.

Eidels, A., Townsend, J. T., & Pomerantz, J. R. (2008). Where similarity beats redundancy: The importance of context, higher order similarity, and response assignment. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1441–1463.

Fifić, M., & Townsend, J. T. (2010). Information-processing alternatives to holistic perception: Identifying the mechanisms of secondary-level holism within a categorization paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1290–1313.

Fitousi, D. (2015). Composite faces are not processed holistically: Evidence from the Garner and redundant target paradigms. *Attention, Perception, & Psychophysics*, *77*, 2037–2060.

Hawkins, R. X. D., Houpt, J. W., Eidels, A., & Townsend, J. T. (2016). Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. *Vision Research*, *126*, 19–33.

Houpt, J. W., Townsend, J. T., & Donkin, C. (2014). A new perspective on visual word processing efficiency. *Acta Psychologica*, *145*, 118–127.

Little, D. R., Nosofsky, R. M., Donkin, C., & Denton, S. (2013). Logical rules and the classification of integral-dimension stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 801–820.

Moneer, S., Wang, T., & Little, D. R. (2016). The processing architectures of whole-object features: A logical-rules approach. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1443–1465.

Wenger, M. J., & Townsend, J. T. (2006). On the costs and benefits of faces and words: Process characteristics of feature search in highly meaningful stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 755–779.

Perceptual Detection

Ben-David, B. M., & Algom, D. (2009). Species of redundancy in visual target detection. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 958–976.

Donkin, C., Little, D. R., & Houpt, J. W. (2014). Assessing the speed–accuracy trade-off effect on the capacity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1183–1202.

Eidels, A., Townsend, J. T., Hughes, H. C., & Perry, L. A. (2015). Evaluating perceptual integration: Uniting response-time- and accuracy-based methodologies. *Attention, Perception, & Psychophysics*, *77*, 659–680.

Lentz, J. J., He, Y., & Townsend, J. T. (2014). A new perspective on binaural integration using response time methodology: Super capacity revealed in conditions of binaural masking release. *Frontiers in Human Neuroscience*, *8*, 641.

Patching, G. R., Englund, M. P., & Hellstrom, A. (2012). Time- and space-order effects in timed discrimination of brightness and size of paired visual stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 915–940.

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.

Yang, C.-T. (2011). Relative saliency in change signals affects perceptual comparison and decision processes in change detection. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1708–1728.

Yang, C.-T., Chang, T.-Y., & Wu, C. J. (2013). Relative change probability affects the decision process of detecting multiple feature changes. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 1365–1385.

Yang, C.-T., Hsu, Y. F., Huang, H. Y., & Yeh, Y. Y. (2011). Relative salience affects the process of detecting changes in orientation and luminance. *Acta Psychologica*, *138*, 377–389.

Temporal Order Processing

Miller, J., Kuhlwein, E., & Ulrich, R. (2004). Effects of redundant visual stimuli on temporal order judgments. *Perception & Psychophysics*, *66*, 563–573.

Visual Search/Visual Attention

Busey, T. A., & Townsend, J. T. (2001). Independent sampling vs interitem dependencies in whole report processing: Contributions of processing architecture and variable attention. *Journal of Mathematical Psychology*, *45*, 283–323.

Fifić, M., Townsend, J. T., & Eidels, A. (2008). Studying visual search using systems factorial methodology with target-distractor similarity as the factor. *Perception & Psychophysics*, *70*, 583–603.

Godwin, H. J., Walenchok, S. C., Houpt, J. W., Hout, M. C., & Goldinger, S. D. (2015). Faster than the speed of rejection: Object identification processes during visual search for multiple targets. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 1007–1020.

Howe, P. D. L., & Ferguson, A. (2015). The identity-location binding problem. *Cognitive Science*, *39*(7), 1622–1645.

McCarley, J. S., Mounts, J. R. W., & Kramer, A. F. (2007). Spatially mediated capacity limits in attentive visual perception. *Acta Psychologica*, *126*, 98–119.

Sung, K. (2008). Serial and parallel attentive visual searches: Evidence from cumulative distribution functions of response times. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1372–1388.

Yamani, Y., McCarley, J. S., Mounts, J. R. W., & Kramer, A. F. (2013). Spatial interference between attended items engenders serial visual processing. *Attention, Perception, & Psychophysics*, *75*, 229–243.

Yang, C.-T., Little, D. R., & Hsu, C. C. (2014). The influence of cueing on attentional focus in perceptual decision making. *Attention, Perception, & Psychophysics*, *76*, 2256–2275.

Zehetleitner, M., Krummenacher, J., & Muller (2009). The detection of feature singletons

defined in two dimensions is based on salience summation, rather than on serial exhaustive or interactive race architectures. *Attention, Perception, & Psychophysics*, *71*, 1739–1759.

Working Memory/Cognitive Load

Chang, T.-Y., & Yang, C.-T. (2014). Individual differences in Zhong-Yong tendency and processing capacity. *Frontiers in Psychology*, *5*, 1316.

Fitousi, D., & Wenger, M. J. (2011). Processing capacity under perceptual and cognitive load: A closer look at load theory. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 781–798.

Heathcote, A., Coleman, J. R., Eidels, A., Watson, J. M., Houpt, J. W., & Strayer, D. L. (2015). Working memory's workload capacity. *Memory & Cognition*, *43*, 973–989.

Schweickert, R., Fortin, C., & Sung, K. (2007). Concurrent visual search and time reproduction with cross-talk. *Journal of Mathematical Psychology*, *51*, 99–121.

Yu, J.-C., Chang, T.-Y., & Yang, C.-T. (2014). Individual differences in working memory capacity and workload capacity. *Frontiers in Psychology*, *5*, 1465.