Information Processing Architectures: Fundamental Issues

James T Townsend and Haiyuan Yang, Indiana University, Bloomington, IN, USA
Trisha Van Zandt, Ohio State University, Columbus, OH, USA
© 2015 Elsevier Ltd. All rights reserved.

Abstract

This chapter starts with a brief history of information processing architectures, emphasizing a traditional experimental paradigm, the additive factors method, and the classic problem of model mimicry. Several solid approaches to identifying mental architecture are introduced with a discussion of the necessary assumptions for these tests. Other fundamental issues attached to information processing systems are also presented. A brief discussion with regard to challenges for the future concludes the chapter.

In the late nineteenth century, budding experimental psychology began to investigate perceptual and mental operations that would be considered part of the contemporary purview of cognition. Such interests reappeared in the 1960s in the emerging information processing approach to elementary cognition. Over the last several decades of the twentieth century, considerable progress has been made in this tradition through the combination of mathematical modeling and experimental design, an approach becoming known as meta-modelling. This section reviews a number of selected findings within this approach, which is relatively novel in the social sciences. One benefit of the strategy is that potential dilemmas involving model mimicking, where models based on strikingly different principles can imitate each other's predictions, are brought to light. In particular, problems and experimental solutions associated with testing among parallel (simultaneous) versus serial (one-at-a-time) processing, stopping rules (logical basis for cessation of processing), limited versus unlimited versus Supercapacity (effects of increasing workload on processing speed), and dependence (issue of stochastic independence of channels or events) relationships are described. The foregoing issues are fundamental in almost all information processing situations since virtually any operating system must take them into account. The section concludes by pointing to future challenges within this approach.

The term 'information processing architecture' refers broadly to the arrangement of mental subsystems that are hypothesized to be active in the performance of one or more psychological tasks. For instance, the simplest, most prototypical, and opposed types of architectures are serial (one-at-a-time) versus parallel (simultaneous) arrangement of two or more separate subsystems or processes. More complex arrangements are mentioned below (see Additive Factor Models). Further, there are a number of other aspects of perceptual and cognitive processing that are often included under the 'architecture' label including the questions of the basis on which cognitive processes will cease, various kinds of independence and dependence, and processing capacity. These will be described and discussed.

An early researcher especially pertinent to this review was Donders, a nineteenth-century Dutch physiologist. Donders believed that he could uncover the durations taken by various thought processes through his method of subtraction. The method of subtraction was based on the idea that complicated mental activities were compounded in a simple sequential fashion from less complex parts. Let mean response time be written as RT and response time in general as simply RT. Then, for instance, the scientist might engage the subject in a task requiring both perception and decision and compare RT from that condition with RT from a task requiring only perception. The difference in RTs was interpreted as the mean duration of the decision operation.

The issues selected for review here seem elemental in the following sense. The construction of almost any system intended to carry out the processing (e.g., detection, search, comparison, recognition, recall, analysis of various kinds, and so on) of a finite number of tasks or items would have to make decisions on each of the studied issues. In addition, they are somewhat unique in having been subjected to quite general theoretical, quantitative, and methodological analysis, perhaps more than any other such concepts in the field. The issues will be introduced in the context of a popular experimental paradigm and then discussed in more detail.

An Experimental Paradigm and the Major Issues

The concepts to be defined below have figured prominently in studies on short-term memory and visual display search and we shall employ that type of paradigm for illustration. Considerable impetus was given to the information processing approach by several pioneering studies in the 1960s focusing on short-term memory and visual processing, using RT as the dependent variable (Atkinson et al., 1969; Egeth, 1966; Sternberg, 1966). In the roughly 45 years of the interim, scores and perhaps hundreds of studies in memory and visual search have been carried out. We focus on Sternberg's (1966) short-term memory search paradigm for illustration. Short-term memory is specified operationally by the tasks requiring the retention of a small number of items for anywhere from a few seconds to several minutes. In this paradigm, a varying number (less than or equal to the number that can be maintained in short-term memory without error) of items, for instance, randomly arranged letters, is presented to the experimental participant. This is called the memory set. Then, a few seconds later, that participant is presented a so-called probe item and the task is...


2 Information Processing Architectures: Fundamental Issues

Another critical concept is that of workload capacity (or 'capacity' for short if the meaning is clear), which refers to how processing times are affected by the number of things to be worked on (Townsend and Asby, 1983). This is most easily illustrated with a version of parallel processing where the system slows down when the number of items that are being processed increases (i.e., limited capacity). However, limitations in capacity that are indirect, even with serial processing, can be conceived. For instance, a serial processor might speed up as it goes through the items, due to warm-up effects, or slow down due to inertial factors or fatiguing of the processor. Even though capacity and independence are logically separate notions, they can interact. For instance, an important type of parallel system, one that can mimic serial processing, assumes that as the processing of each item is completed, its processing capacity is reallocated to remaining items (Townsend and Asby, 1983). This obviously affects the overall RT, but also creates a positive dependence among the item processing times. Townsend and Wenger (2004b) build a theory and associated methodology featuring the capacity theme.

Yet another important notion is that of stopping rule (e.g., Sternberg, 1966, Townsend, 1974). Depending on the task, it may or may not be necessary for the participant to process all of the items in order to make a correct response. In the memory search paradigm, if the probe item is present and located in the current stimulus set, the processing can cease at that instant, without finishing the remaining items. This possibility is known as self-termination. However, since short-term memory search consumes only a few hundred milliseconds it is possible that the system nevertheless completes all items. This event is called exhaustive processing. On probe-absent trials, it is necessary to process all of the memory items in order to be sure of correctly making a 'no' response, that is, exhaustive processing must occur. In some experimental designs, all the items are probes. These need not be physically identical. For instance, the task might be to say 'yes' if any of the items is a vowel and a target-present trial could contain all vowels. This latter case permits the possibility of first-terminating or minimum time processing. Of course, it is an empirical question as to whether any kind of self-termination can actually take place in high-speed perceptual or cognitive operations, and must be addressed experimentally in each case.

Space constrains the scope of mathematical detail in the present article, but we provide some fundamental, if simplified, depiction. Although the serial and parallel classes of models both contain an infinite class of possibilities, the serial notion has traditionally been attached to a particular serial model that assumes identical processing time random variables on each item, independent of the number of items in memory (i.e., the load) and of the order of processing the items. It is also often assumed that the individual processing times are themselves mutually stochastically independent. We call this the standard serial model. Let the density of processing time for each item be designated \( f(t) \) and that for the independent residual processing time be \( g(r) \), where \( T \) and \( R \) are the respective random variables for \( t \) and \( r \). Then the density on a probe-absent trial, \( p(RT = t) \), or \( p(t) \) for short, is just the mathematical convolution of the \( n \) processing densities and
the residual time density: $p_i(\text{RT}) = f_i * f_j * \ldots * f_n$, where each $f_i, i = 1, \ldots, n$, is a replica of. The expected or average theoretical RT is $E(\text{RT}) = n * E(T) + E(R)$. Notice that processing is exhaustive, as designated by EX in the left-hand side, in this case. In the case of self-terminating (ST) processing on a probe-present trial, the number of items finished before the probe is found is itself a random variable. Under usual conditions, the probability that the probe will be found in the $i$-th processing position is just $1/n$, so the average density for this case is just $E[p_i(\text{ST})] = (1/n) \sum_{i=1}^{n} f_i * f_j * \ldots * f_n$. Similarly, the expected RT, with RT as the sample statistic, is just $E[\text{RT}] = E(T) + E(R) = (n+1)/2$. $E(T)$ is the processing time, that is, $PT = RT - R$. Then $E[\text{RT}] = \int_{c}^{d} \left[1 - G(t)\right]dt$, where $G(t)$ is the cumulative distribution function associated with $g$. It is straightforward to show from this formula that mean RT is indeed an increasing, concave function of $n$ (Townsend and Ashby, 1983, p. 92). Note that $g(t)$ is invariant across values of $g$. Hence, many investigators use a very restrictive and typically unrealistic criterion for parallelism when demand flat rather than increasing functions RT of $n$, even under exhaustive processing conditions.

It can also be observed that if mean exhaustive RTs are flat functions of $n$, then mean ST (again, a single position exists in the memory list that contains the probe) times would actually be decreasing, within the same model – a strong prediction which apparently has never been checked in studies using this logic. Thus, in the rare cases where flat exhaustive RT functions are actually found with exhaustive processing, the implications are quite strong and in favor of very supercapacity parallel processing. Models that can make such predictions are considered in Townsend and Ashby (1983, Chapter 4).

In contrast, in the case of ordinary ST processing when a single position contains the probe, standard parallel models do predict a flat $RT$ function. This can be readily intuitively, since only the channel processes the probe matters in ST processing, and it is independent of all the rest (and is of unlimited capacity into the bargain). Finally in this model, mean first-terminating times should decrease with $n$.

## Important Related Research Topics

Although the issues considered here are applicable to virtually any information processing system or task, there are some classical or contemporary subject areas for which they are particularly germane. When applied within the confines of a single subprocess, such as within a memory or display list, the concept of attention is evoked. How attention is deployed, when it ceases, whether its application is independent across the various items, and its strength as a function of the workload, all call upon the critical processing issues introduced above.

The topic of automatic processing traditionally discussed within the realm of attention, likewise overlaps our present issues (e.g., Schneider and Shiffrin, 1977). The most obvious correlate within our processing issues is that of capacity. Although writers are sometimes rather vague concerning detailed quantitative accounts of what automaticity means with regard to workload capacity, it seems apparent that parallel processing is definitely mandated. In addition, we have proposed that capacity should at least be at the unlimited level. That is, each subsidiary task should see its operations proceed at the same speed as if it were functioning alone. These and related matters are discussed further in Neufeld et al. (2007).

Another traditional theme, but also seen resurfacing, is that of perceptual and cognitive holism, a topic of long standing in philosophy and in the twentieth century, of Gestalt psychology. Although much has been learned about psychological holism, there has been little accomplished with regard to definitions, theory, and experimental investigation of information processing characterizations of dynamic holistic operations. In that spirit, a set of working axioms which portray holistic perception in terms of the present issues have been proposed and developed (e.g., Wenger and Townsend, 2001; Wenger and Ingvallson, 2003; Fill and Townsend, 2010). The accompanying essentials of holistic perception can be informally expressed as involving highly interactive, supercapacity, exhaustive, parallel processing.

## The Mimicking Dilemma of Serial versus Parallel Processing

As mentioned earlier, one of the driving forces behind mathematical metamodeling in this area was the discovery that mathematical representations of diagnostically opposed psychological principles could nevertheless sometimes mimic one another, even to the point of mathematical equivalence of their defining probability laws (Murdock, 1971; Townsend, 1971; Townsend and Ashby, 1983). A historical account of this dilemma and its resolution is offered by Townsend et al. (2011). Hence, we will take some time here to outline the state of the art with regard to such questions within the parallel versus serial question.
Consider a model for the processing times of $p$ items under fixed experimental conditions. With regard to the parallel-serial issue, suppose no assumptions are made other than the probability mixture of generalized convolutions in the case of seriality and joint distributions on processing times in the case of parallelism. Then the parallel and serial classes of models are equivalent, in the sense that mappings can be provided that homeomorphically (this is mathematical jargon for a one-to-one, and continuous to-and-fro function relating the two types of models) carry the serial probability distribution into the parallel probability distribution and vice versa (Townsend and Ashby, 1983).

Nevertheless, over time, accumulating theoretical results have demonstrated that if the scientist is willing to make further restrictive, but sometimes still very general (and reasonable), assumptions about the models and/or more complex experimental designs are utilized, the parallel-serial issue can be decided. For instance, certain rather fundamental differences between serial and parallel processing can be explored in experimental methods designed to exploit those differences (Townsend and Ashby, 1983; Townsend and Wenger, 2004a).

One of the most promising and general approaches to identifying mental architecture (i.e., serial versus parallel processing) is that based on the notion of selective influence of experimental factors, a notion first employed in tests of strict seriality by Sternberg (1969) in his well-known additive factors method (see Additive Factors Models). All factorial methodologies, like the original Sternberg strategy, depend on the selective influence assumption, namely, that distinct experimental factors affect distinct processing components (i.e., subsystems), the assumption of selective influence. It can be assumed that $\text{RT}(X + \Delta X, Y)$ refers to the case where the $X$ factor has prolonged RT but $Y$ is at base level, and so on. Basically, the fundamental statistic for the original method and for most extensions was the mean interaction contrast (MIC). The MIC is defined as

$$\text{MIC} = \frac{\text{RT}(X + \Delta X, Y + \Delta Y) - \text{RT}(X + \Delta X, Y) - \text{RT}(X, Y + \Delta Y) + \text{RT}(X, Y)}{\text{RT}(X, Y)}.$$ 

Schweickert (1978) in his latent mental network theory contributed the first major extension of the additive factors method involving more complex architectures under the assumption of selective influence. Townsend and Ashby (1983) found that the MIC distinguished parallel and serial stochastic models when selective influence was assumed and Schweickert and Townsend (1989) produced general theorems for Schweickert’s latent networks, within a stochastic setting, assuming exhaustive processing (see Additive Factor Models; 4:3069).

Although the early theorems in all this work were accomplished in the context of exhaustive processing, analogous results can be found in the case of ST and first-terminating processing times (e.g., Townsend and Nozawa, 1995; Townsend and Wenger, 2004a). Because Sternberg’s original ideas inherent in his additive factors method have been extended in so many new directions, it has been suggested that the general approach be referred to as systems factorial technology (Townsend and Thomas, 1994). For instance, one novel strategy has been to enlist entire RT distributions in providing more powerful tests of parallel versus serial processing or other related architectures (Dzhafarov and Schweickert, 1995; Roberts and Sternberg, 1993; Townsend, 1990; Townsend and Ashby, 1983; see also Balakrishnan, 1994). For instance, in analogy to the MIC, the scientist can form a contrast function composed of the double difference (corresponding to the double difference in mean RTs in the usual case) of cumulative distribution functions. This new statistical function turns out to be very helpful in assaying mental architecture (Townsend and Nozawa, 1993). Another example of the use of entire distributions will be reviewed below.

The question may be raised as to whether processing issues other than parallel versus serial processing also suffer significant problems in identification within the search or related paradigms. With regard to the stopping rule, mathematical investigations have shown, somewhat ironically, that the same types of memory search data that were incapable of deciding the parallel-serial question could in many instances prove that processing was ST rather than exhaustive. Interestingly, it is more difficult to prove that processing is exhaustive in that ST models can mimic exhaustive processing but not vice versa (Townsend and Colonius, 1997; Van Zandt and Townsend, 1993).

Moreover, recent theoretical and empirical discoveries utilizing the entire RT distributions rather than means provide much strengthened tests of architecture (parallel versus serial) and in addition allow one to firmly distinguish stopping rules at the same time. In fact, within systems factorial technology and assuming selective influence at the distributional level, Townsend and Nozawa 1995 (see also Townsend and Wenger, 2004a) proposed an experimental design that they called the double factorial paradigm in which investigators can test architecture and stopping rule as well as capacity within the same block of experimental trials.

Additionally, the important matter of workload capacity seems to possess little in the way of mimicking predicaments. Whether capacity increases, decreases, or is invariant, in the face of workload alterations is immediately captured by the current statistical measures. The most difficult issue of the ones under discussion is that of stochastic dependence, even though it plays a vital role in processing systems. Interestingly, at this point in time, dependence is more readily and more directly assessed within experimental
Another evident opportunity for development is the exten-

sion of factorial methods in general, and systems factorial
technology in particular, to other dependent variables than
response times. The most palatable extension would be to
accuracy or accuracy and response times. Another interesting and
topical subject matter would be confidence ratings. In this regard,
it has been pointed out that it constitutes a grave error to simply
assume that predictions for response times will also apply to
other observable variables (Townsend, 1984). Theorems that are
appropriate for specific observables within an information
processing milieu must be proven and tested in the experi-
nmental crucible.

As a final exemplary vein that is offered is that of imple-
menting the concept of a theoretical and methodological siev-
e. Within such a sieve, the investigator first establishes the major
processing characteristics of information processing in a cogni-
tive task, and then proceeds to propose and test more computa-
tionally detailed accounts that are in accord with the earlier
findings. The methodological sieve is related to, but is more
general than, the concept of using inference put forth by Platt
(1964) (e.g., Massaro, 1998).

Bibliography


Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-

Egeth, H., 1966. Parallel versus serial processes in multidimensional stimulus discrim-
6 Information Processing Architectures: Fundamental Issues