

**Independent Sampling vs. Inter-Item
Dependencies In Whole Report Processing: Contributions of Processing
Architecture And Variable Attention**

Thomas A. Busey

and

James T. Townsend

Indiana University

Please send correspondence to:

Thomas A. Busey
Department of Psychology
Indiana University
Bloomington, IN 47405

email: busey@indiana.edu

In Press, Journal of Mathematical Psychology

Abstract

All current models of visual whole report processing assume perceptual independence among the displayed items in which the perceptual processing of individual items is not affected by other items in the display. However, several models proposed by Townsend (1981), Shibuya & Bundesen (1988) and Bundesen (1990) contain post-perceptual buffers that must predict negative dependencies. The perceptual independence assumption forms what we term the Modal Model Class. A recent example of a model that assumes perceptual independence is the Independent Sampling Model of Loftus, Busey & Senders (1993). The fundamental independence assumption has only been directly tested once before (Townsend, 1981) where tests revealed no dependencies except those produced by guessing. The present study provides new tests of the independence assumption, and finds significant positive dependence, which is inconsistent with most extant models of whole report. Poisson models do predict a positive dependence and we develop a succinctly parameterized version, the Weighted Path Poisson Model, that allows the finishing order to be a weighting probabilistic mechanism. However, it does not predict the data quite as well as a new model, the Variable Attention Model, that allows independence within trials (unlike the Poisson models). This model assumes that attention (or potentially, other aspects such as signal quality) varies widely across trials, thus predicting an overall positive dependence. Intuitions for and against the competing models are discussed. In addition, we show, through mimicking formulae, that models that contain the proper qualitative type of dependence structure can be cast in either serial or parallel form.

The advent of the information processing approach in the late 1950's and 1960's led not only to the isolation and analysis of new processing subsystems and studies of their action and interaction, but also to the development of mathematical models of a variety of these systems and mechanisms. One of the earliest and most prominent to receive attention was the short-term memory paradigm known as "whole report" or "full report", in which the subject was instructed to report verbally as many unrelated items as she could from a visual display. Sperling's (1960) classic investigations combined his partial report paradigm with various manipulations and analyses on that and whole report data. The results separated the early visual information storage, to become known as the icon, from subsequent short-term memory and rehearsal mechanisms. The whole report paradigm has continued to be important in its own right and models of the initial perceptual processing stages have been concatenated with other perceptual and cognitive processes.

Early general opinion seemed to be that perception of the displayed items in whole report took place in a serial manner, undoubtedly suggested by the linear-appearing (at least in a limited range) average-number-correct functions of exposure duration (Sperling, 1960). However, most later investigators came to favor a parallel interpretation in the perceptual stages of processing (e.g. Sperling, 1967). A notable early example of parallel models of whole report was Rumelhart's Multicomponent Model (1970) which accommodated many of the basic whole report phenomena but also was able to predict a number of aspects of partial report. Within the whole report paradigm, the Multicomponent Model assumed independence of item processing. However, the assumption of no inter-item dependencies was not tested at the time.

Since that time, a number of other models of whole report behavior have appeared. All quantified models, with the exception of several tested by Townsend (1981), assume that the separately displayed items were processed independently by the early stages of visual processing. Further, all except the Multicomponent Model also assume that probability correct at any signal location is proportional to an exponential function or that probability is itself an exponential distribution. The exact meaning of this distinction will be made explicit below. Examples of independent whole report models include those by Rumelhart (1970), Townsend (1981), and Loftus, Busey & Senders (1993). The perceptual processing stage of a model by Shibuya & Bundesen (1988) assumes independence, but it is followed by a memory stage, which forces a negative dependence, as will be seen later. Because of their shared characteristics, we refer to all models based on independence and as members of the Modal Model Class. They are not the same model, because each was based on somewhat different additional assumptions. Because they are not

the same model, and the interpretation of analogous parameters was typically different across investigators, we employ the original notation when it is necessary to refer to them. That should also make it easier for readers to consult the original manuscripts.

The major goal of the present study is to test the assumption of inter-item independence and use it to discriminate between candidate models. We will demonstrate, using extant data, the existence of inter-item dependencies, and use the structure of the dependencies to develop two new models that can closely approximate the observed dependency structure. These dependencies are important, because they are positive, and virtually every other model of whole report either predicts no inter-item dependencies or negative inter-item dependencies. Thus our data bears on an entire class of models that share the independence assumption, and often assume an exponential distribution assumption as well.

It is of critical importance to describe the relation between independence and capacity, since certain capacity assumptions can produce inter-item dependencies even though the individual items are *perceived* independently within any given trial. For example, one form of capacity is a fixed size memory that does not have storage room for all of the displayed items. In this case, if an item is acquired and placed into memory, it leaves less room for other items and thus produces a negative dependency overall, even if the items are perceptually processed independently on any particular trial (see the Fixed Sample Size Model, Townsend, 1981 and the Fixed Capacity Independent Race Model of Shibuya & Bundesen, 1988). However, other forms of capacity limitation affect only the processing rates of individual items, and do not necessarily produce inter-item dependencies.

A general taxonomy of the models we consider is shown in Figure 1. In characterizing the models, we consider three components. The first is whether the stimulus processing is described by an exponential growth function or by an exponential (or other) probability distribution. The predictions that result in the case of an exponential growth function vs. an exponential probability distribution are very similar and sometimes identical, even though the underlying assumptions differ. Second, we consider whether the rate of processing of individual items in the display is affected by the number of items in the display. In the visual search literature this is commonly known as set size effects. If the rate of processing of *individual items* decreases as the number of to-be-reported items increases, we characterize the models as limited capacity. Note that although capacity is limited, the processing of individual items can remain independent, and thus a limited capacity assumption at this stage of processing does not necessarily impose inter-item

the same model, and the interpretation of analogous parameters was typically different across investigators, we employ the original notation when it is necessary to refer to them. That should also make it easier for readers to consult the original manuscripts.

The major goal of the present study is to test the assumption of inter-item independence and use it to discriminate between candidate models. We will demonstrate, using extant data, the existence of inter-item dependencies, and use the structure of the dependencies to develop two new models that can closely approximate the observed dependency structure. These dependencies are important, because they are positive, and virtually every other model of whole report either predicts no inter-item dependencies or negative inter-item dependencies. Thus our data bears on an entire class of models that share the independence assumption, and often assume an exponential distribution assumption as well.

It is of critical importance to describe the relation between independence and capacity, since certain capacity assumptions can produce inter-item dependencies even though the individual items are *perceived* independently within any given trial. For example, one form of capacity is a fixed size memory that does not have storage room for all of the displayed items. In this case, if an item is acquired and placed into memory, it leaves less room for other items and thus produces a negative dependency overall, even if the items are perceptually processed independently on any particular trial (see the Fixed Sample Size Model, Townsend, 1981 and the Fixed Capacity Independent Race Model of Shibuya & Bundesen, 1988). However, other forms of capacity limitation affect only the processing rates of individual items, and do not necessarily produce inter-item dependencies.

A general taxonomy of the models we consider is shown in Figure 1. In characterizing the models, we consider three components. The first is whether the stimulus processing is described by an exponential growth function or by an exponential (or other) probability distribution. The predictions that result in the case of an exponential growth function vs. an exponential probability distribution are very similar and sometimes identical, even though the underlying assumptions differ. Second, we consider whether the rate of processing of individual items in the display is affected by the number of items in the display. In the visual search literature this is commonly known as set size effects. If the rate of processing of *individual items* decreases as the number of to-be-reported items increases, we characterize the models as limited capacity. Note that although capacity is limited, the processing of individual items can remain independent, and thus a limited capacity assumption at this stage of processing does not necessarily impose inter-item

that the latter function decreases as n grows, for any time t . Then clearly, the probability of any specific pattern of corrects and errors is given by the binomial distribution, which obviously predicts independence, even though as n grows, the individual item probability correct decreases, for any fixed arbitrary time t . The primary difference between the Rumelhart and Townsend models with respect to capacity effects will be specified quantitatively below. Additional information on the relation between capacity and inter-time dependencies, particularly with respect to the models analyzed in this study, is found in Appendix A.

An exception to these models is the Fixed-Path Poisson Model, which will be discussed later in more detail. In its serial interpretation on which we focus here, it has unlimited capacity at the level of an individual item, but all of that capacity is assigned to process one item at a time in a serial fashion. Capacity is unlimited here at the level of an individual item because it does not depend upon the total number of items in the display; each one is processed until all are complete or a mask terminates processing. As we will see, this particular model predicts positive inter-item dependencies.

To summarize, the whole report models are either limited or unlimited capacity at the level of the individual item, and may contain a bound that limits performance to less than 1.0. However, neither mechanism necessarily by itself or together introduces inter-item dependencies. A fixed memory buffer with a limited number of entries will produce negative dependencies, while our serial Poisson models that work on one item at a time will produce positive dependencies, as described in a later section.

Formulations of Extant Models of Whole Report

The previous section provided an overview of the extant models of whole report. In this section we briefly describe the major assumptions underlying the Bounded Performance Model (Townsend, 1981), the Fixed Capacity Independent Race Model of Shibuya & Bundesen (1988), and the Independent Sampling Model (ISM, Loftus et al (1993), which will demonstrate whether (and how) inter-item dependencies are predicted for each model.

The Bounded Performance Model posits that growth of the available stimulus information $I(t)$ in any given item (e.g., letter or digit) proceeds according to the elementary differential equation $dI/dt = [I(\infty) - I(t)]V$, with the solution, $I(t) = I(\infty)\{1 - \exp[-(t-t_0)V]\}$, where $I(\infty) = \lim_{t \rightarrow \infty} I(t)$, i.e. the asymptotic level of information. "V" is the rate of growth, which is assumed to depend on stimulus parameters such as

intensity, and the subject's sensitivity, and t_0 is the 'lift-off' time at which information begins to be accumulated. That basic process occurs independently on each stimulus location. Next, the Bounded Performance Model makes the assumption that the subject has available an amount of capacity, ζ . This capacity is allocated, perhaps unevenly across the various displayed items, giving rise to non-uniform serial position effects. The results of experiments that vary the display size (i.e. processing load), find that capacity ζ does not expand to accommodate the additional items, demonstrating that processing is limited capacity, even though processing is stochastically independent for any given display size. Typically we find that capacity is limited for displays larger than $n=5-6$ items, in the sense that average number correct is asymptotically less than n for very long exposure durations. For instance, Townsend (1981) compared the Bounded Performance Model to Rumelhart's Multicomponent Model, and found that performance in a five-consonant whole-report design in the Townsend study was strongly bounded even as time became large, up to 250 msec. Thus, asymptotically, the average number of correct reports was bounded below 5, the number of stimuli displayed. This bound could not be captured by the Multicomponent Model, which as noted, predicts that performance would ultimately increase to perfection as duration increased. The Bounded Performance Model demonstrates how this can be done: the model includes the assumption that the observer allocates some amount of capacity, a_i , to each item i , such that $\sum_i a_i = 1$. Accuracy is then governed by

$$P(\text{cor on loc. } i \text{ at time } t \geq t_0) = P(C_i, t) = \text{Min} \left[\frac{I(t)a_i}{I(\infty)\zeta}, 1 \right]$$

$$= \text{Min} \left[a_i \zeta (1 - e^{-(t-t_0)/V}), 1 \right]$$

Thus, probability correct is basically the product of the allocated capacity and the available stimulus information. Note that the expected number correct equals

$$E(\text{No. Cor.}, t) = \sum_i P(C_i, t) = \left[1 - e^{-(t-t_0)/V} \right] \zeta \text{ with } \lim_{t \rightarrow \infty} E(\text{No. Cor.}, t) = \zeta \text{ if } \zeta < n .$$

However, the data

discussed in the present article are derived from experiments with a fixed load of 4 digits so we may dispense with explicit capacity notation in the independence models. It is important to observe that performance does not go to perfection because probability of correct perception is proportional to an exponential function (not the exponential distribution here) that is bounded. Rumelhart's (1970)

Multicomponent Model is also independent and of fixed capacity (for whole-report processing), but in that model $P(C_i, t)$ approaches 1 as t gets large, because the explicit formula for this function is an actual distribution that goes to 1 as time gets large. The distribution is gamma, even though an exponential, or any, distribution would yield comparable qualitative properties. Again, observe that for a given t , Rumelhart's probability correct function decreases monotonically with n , just like the Bounded Performance Model, because it is of limited capacity.

Another more recent model included in the Modal Model Class is the Fixed Capacity Independent Race Model of Shibuya & Bundesen (1988), which is based on independent exponential random variables. This model was primarily developed for experiments where a variable number of targets-to-be-reported are randomly placed among a variable number of distractors, which are to be ignored. Thus, their primary focus was on a form of partial report but, a sub-condition contained what was, in essence, a whole report task. Although we cannot enter into details about the partial report aspects, several aspects are worth noting, since the model performed well with their data. The sampling process begins some fixed time after the display starts (analogous to t_o in Townsend's model and to L [see below] in the Loftus, et al. model). The front end sampling was comprised of a limited amount of capacity spread over the presented items, thus rendering this part a form of the C/n capacity class of models mentioned above. Then, each item is processed exponentially with a rate proportional to C/n (and depending on an earlier selection criterion). The sampled items are placed in a fixed capacity short-term store. Processing terminates either when the size of the short-term store is equaled, or when a fixed time after the mask appears, has elapsed. The upshot is that this model has limited capacity, but independent perceptual processing followed by the fixed-size memory, which will tend to cause negative dependencies. The fixed duration of processing allowed, until terminated by the mask, will not affect dependencies but can form an upper bound to accuracy, as found in the data from Townsend (1981).

A defining feature of this model is that the completion times for displayed items follow exponential distributions which will produce asymptotic performance approaching perfection as t gets large. Completed items are delivered to a fixed storage capacity memory, apparently identical in form to that proposed by Townsend's (1981) Fixed Sample Size Model, from which they are read out by the subject as they respond. The existence of a fixed capacity memory must predict negative inter-item dependencies, since if one item is acquired, there will be less room for other items in memory. A later version

(Bundesen, 1990), allows a mask to terminate processing and potentially lead to an (imperfect) upper bound on performance, even without the subsequent fixed capacity memory. However, this model still predicts negative dependencies, as long as the input to the storage buffer exceeds its fixed capacity. The dependence predictions were not tested explicitly, but the overall fits apparently tested that assumption implicitly.

In sum, all models (post-1981) of whole report processing with the exception of the Poisson model (Townsend, 1981), predict either no inter-item dependencies or negative inter-item dependencies due to fixed-size memory buffers. That most models do not make predictions for inter-item dependencies may not come as a surprise; it may be surmised that most models were originally designed to handle marginal data only. Independence of various types has been the initial default assumption in philosophy and science at least since the time of the Greeks. However, the data from whole report tasks intrinsically includes information about inter-item dependencies, and thus offers clues to a deeper understanding of underlying process mechanisms. If current models fail to capture the dependency structure, an obvious step is to extend them, or propose new models that accurately account for the inter-item dependencies.

The member of the Modal Model Class on which we focus in this paper was developed by Loftus and his colleagues in several recent articles (Loftus, Duncan, & Gehrig, 1992; Loftus & Busey, 1992; Loftus & Ruthruff, 1994; Loftus, Busey & Senders 1993; Busey & Loftus, 1994). They proposed the Independent Sampling Model of visual information processing to account for performance in a four-item whole report task. Recently, that model has been expanded to include linear filtering mechanisms that address the nature of the initial time-varying sensory representation from which information is extracted. However, it is the model's original form in which we are most interested, and below we describe its major assumptions.

The Independent Sampling Model (also termed the Random Sampling Model) describes information processing as the acquisition of stimulus features through a process that randomly samples features from the stimulus with replacement, at a rate that remains constant through the stimulus display duration. The favored interpretation is one in which the stimulus information is acquired according to an exponential growth function. The time-invariant stimulus acquisition rate represents the raw feature-sampling rate. Because features are sampled with replacement, the rate of acquisition of new features is proportional to one minus the proportion of already sampled features. These acquired features subsequently form the basis

for the sensory representation. Because sampling occurs with replacement from the features in the display, the result is a theory that predicts that the rate of acquisition of new features is exponentially distributed. To this effect, let $\mathbf{X} = \{0,1\}$ be a random variable designating incorrect ($\mathbf{X}=0$) or correct ($\mathbf{X}=1$). The random variable notation will be useful for later formulas. This conceptualization, with an additional assumption relating the proportion of acquired features to proportion of correctly-recalled items, relates stimulus duration to performance on each item via the expression:

$$P(\text{Correct}) = P(C) = P(\mathbf{X}=1) = (1.0 - e^{-(d-L)(1/c)}) + g(e^{-(d-L)(1/c)}) \quad \text{Eq. 1}$$

where $P(C)$ represents proportion-correct performance, g represents the guessing rate, d represents stimulus duration, $1/c$ represents the raw stimulus rate, L (for "Liftoff") represents the pre-processing delay (termed t_0 in other models), and the exponential captures the sampling-with-replacement property of the independent sampling model¹. Eq. 1 could be amended to include a bound, as discussed below.

Note that the form of exponential growth here is analogous to Townsend's Bounded Performance Model, not an exponential distribution in the usual sense. The first part of Eq 1 represents the probability of obtaining the digit via the random sampling mechanism, while the second part represents the probability that the item is correctly guessed if it is not obtained from the random sampling mechanism. The guessing rate g can legitimately be set to $1/10$, since in the present study all stimulus items were sampled by the experimenter with replacement from the ten digits. (The Townsend (1981) study sampled from the 20 consonants without replacement which necessitated a more complex guessing structure.) Although the notion of a bound on performance was not an integral part of this model, some provision was permitted for that due to ancillary disturbances, such as eye movements, lapse of attention etc.². Observe that Eq. 1 does not exhibit position effects, but that characteristic is easily relaxed as in Townsend's (1981) model.

¹We will incline toward using a " $P(C_i)$ " type of notation for probability correct in location i for simple formulas, but the random variable expressions will be needed for certain joint probabilities.

²The original Loftus, Busey & Senders (1993) formulation included an asymptote parameter that could be less than 1, and was assumed to represent keypress errors and not memory limitations. However, the estimated asymptotes were very close to 1.0 (.84, .98 and .98 for subjects TB, SS and EF, respectively) and the search space was very flat, such that the variable asymptote increased the r^2 values only marginally (.949 to .951; .998 to .999; and .970 to .995 for the three subjects). Application of the structurally isomorphic Bounded Performance model (but based on somewhat distinct assumptions) found the lower asymptote to be necessary with five-consonant stimuli (Townsend, 1981). However, the asymptote does not have much explanatory power in the Loftus et al (1993) (and in the present data), and the inclusion of the variable asymptote makes several analytic analyses intractable. Thus we have chosen to eliminate it from the current discussion.

Investigation of the dependency structure in a task takes us a step deeper into the architecture of processing mechanisms and increases our understanding of the details of the time-dependent processing characteristics. For instance, continued findings of independence (as in Townsend, 1981) would argue not only for separate independent channels of perceptual processing, but also indicate that subsequent stages (in parallel or in serial) of processing do not seriously compromise that independence³. For example, a fixed capacity memory buffer as in the Shibuya & Bundesen (1988) model or Townsend's Fixed Sample Size model, would introduce negative dependencies even if individual items are processed independently to start with. Thus, an analysis of inter-item dependencies can shed light on the processes that underlie both the acquisition of individual item information as well as the architecture of later translation, memory and output stages. This is accomplished by examining the structure of the dependencies using a variety of statistics that provide an intuition into the structure of the data. Candidate models are then used to produce predictions that contain inter-item dependencies that can be compared to the obtained data using log-likelihood statistical tests.

With these goals in mind, we reanalyze data previously collected by Loftus, Busey & Senders (1993) to see if the independence findings of Townsend (1981) would be replicated, and if not, what sort of relatively simple model might explain the dependencies. In the Loftus et al. (1993) study, subjects viewed briefly-presented stimuli consisting of 4 randomly-chosen digits (chosen with replacement) presented using a slide-projector based laboratory. The digits were low contrast (~7.5%) and viewed under mesopic luminance conditions. The digits were shown for one of 8 stimulus durations ranging from 40 to 200 ms, and were immediately post-masked by a pattern mask. The subject typed in the 4 digits on a keypad, starting from the left-most digit and working to the right. If an item was correct but in the wrong location, it was counted as wrong. If a digit was not perceived, the subject was instructed to guess. The primary dependent measure was the proportion of correctly-recalled digits for each stimulus duration, but, as we will demonstrate below, other statistics can be computed.

Performance in the whole-report task is computed as the number of correctly recalled digits, and averaged within each stimulus duration. These values are plotted against stimulus duration to get a

³Of course it would always be possible, if unlikely, to produce independence through a complex trade-off of negative and positive dependencies.

marginal performance curve. Figure 2 shows data from three observers in the original Loftus, Busey & Senders (1993) Experiment 1. The long-dashed curves represent the Independent Sampling Model prediction for this task, while the short-dashed curve represents the prediction from a particular model based on the Poisson distribution, to be described below. Estimated parameter values for these and other models are found in Table 1.

INSERT FIGURE 2 AND TABLE 1 ABOUT HERE

We should note that attention does not vary across trials in the basic version of the Independent Sampling Model. However, in a subsequent section we add this assumption to the model.

The remainder of the article is organized as follows. Recall that the hallmark of the independence assumption is that there are no inter-item dependencies; that is, the probability of acquiring and reporting one item is independent from the probability of acquiring and reporting some other digit. We demonstrate that the data contain positive dependencies, which are inconsistent with the independent sampling assumption and models that predict negative dependencies, such as the Shibuya & Bundesen model (1988; Bundesen, 1990) or the fixed size buffer model of Townsend (1981). To account for these dependencies, we develop and extend a class of Poisson models that can also account for the marginal data shown in Figure 2. We then demonstrate how one of the Poisson models provides a better account of these data than the Independent Sampling Model. Finally, we show how a variant of the Independent Sampling Model that assumes that attention varies across trials can, by assuming substantial variability in attentional (and/or stimulus quality) states, also account for the positive dependencies. This account, as we will see, relies on apparently independent shifts in attention that seem contrary to the traditional view of attention varying gradually across trials.

Evidence for Inter-Item Dependencies

A first step in our analyses is to determine whether inter-item dependencies exist in the data, as described below. If we find evidence for dependencies, these can be further explored using several statistics described in subsequent sections. Finally, the Independence and Poisson models will be tested using traditional log-likelihood analyses.

Independence Analyses

A non-parametric test of the dependencies between items is described by Kullback (1968), which provides for tests of independence within a single stimulus duration but makes no assumptions about the

growth curve. Kullback (1968) computes an Independence statistic that is similar to a chi-square test, which places the result of each trial in one of 16 cells according to whether the digit in each of the four positions was correct or incorrect. These counts are indexed as x_{ijkl} where i through l index positions 1 through 4 of the 4 digits. For example, suppose the observer produces a result of $\{0, 1, 1, 1\}$. Thus $i = 0$ and $j..l = 1$. For this trial we would increment $x_{0,1,1,1}$ by 1 to reflect this outcome. The independence statistic is:

$$Independence = \sum_{i=0}^1 \sum_{j=0}^1 \sum_{k=0}^1 \sum_{l=0}^1 x_{ijkl} \log \frac{N^3 x_{ijkl}}{x_{i...} x_{.j..} x_{..k.} x_{...l}} \quad \text{Eq. 2}$$

where $x_{i...}$ is the sum of the correct and incorrect digits in positions 2-4 for a given state of i (correct or incorrect). The other terms represent similar marginal statistics for the other positions. N represents the total number of trials in the experiment at a given exposure duration. This statistic is distributed as chi-square with 11 degrees of freedom for the 4 digit stimuli. The critical value for $\alpha = 0.05$ is 19.67.

Such a statistic must be computed separately for different stimulus durations, since dependencies will be produced if trials are mixed from different exposure durations. This has the disadvantage, common among chi-square-type analyses, that a marginal in the denominator of Eq. 2 might be zero, making the term undefined. When this occurred we removed this term from the summation. This has the effect of reducing the Independence statistic from its true value, which makes it more conservative since it must exceed the critical value in order to demonstrate evidence of dependencies.

Table 2 lists the computed independence values for the three observers. Values that exceed significance are listed in bold type.

INSERT TABLE 2 ABOUT HERE

The independence statistics in Table 2 demonstrate significant inter-item dependencies for all three observers. Short stimulus durations produce chance performance and thus no dependencies are to be expected, while Observer EF's performance is near 1.0 for long stimulus durations, making inter-item dependencies impossible. However, at moderate stimulus durations (relative to each observer) we see clear evidence of inter-item dependencies.

A disadvantage of the independence statistic is that it does not directly reveal the direction of the dependencies or give much sense about what is happening to the dependencies as, say, display duration

varies or the number of other items correct increases in the joint probability. Two statistics that partially fulfill this need are described below. These statistics are intended to provide an intuitive summary of the dependency structure, and in a subsequent section we will provide log-likelihood tests that exhaustively test the dependencies in the data. These latter tests will provide the best techniques for discriminating between models.

Alternative Candidate Models

We consider two alternative classes of models that might account for the dependencies found in the previous analyses. The first class, drawn from the Poisson family of distributions, is typically interpreted as a serial mechanism in which attention moves a single processing mechanism in some fixed serial order through the processing locations. Although originally suggested as a serial translator from iconic visual information storage to an acoustic buffer (Townsend, 1981; as suggested qualitatively by Sperling, 1967), we will see it can also accommodate parallel interpretations. A second type of model, the Variable Attention Model, extends the parallel processing nature of the Independent Sampling Model, by assuming that either the processing capacity (as in attentional states) or the quality of the incoming information, varies across trials. At the beginning of a trial, a sample is taken from the distribution on capacity, which then stays fixed throughout that trial. That amount of capacity can then be distributed, perhaps in an uneven manner, across the various item locations. However, it is assumed that the ratios of capacity across the item locations are constant across trials, implying that the sample capacity on that trial simply scales up or down the average capacity at each location. For instance, if the observer is in a high attentional state for a given trial, processing will speed up at all four stimulus locations by an amount that is proportional to their average processing rate.

While these models may be given different architectural interpretations, one way to think about the models is in terms of how attention varies during processing. In the original Independent Sampling Model, attention is fixed across trials, but can be different for the different stimulus locations in order to account for serial position effects. The Poisson models can be thought of as attention moving a single process from one location to the next in a serial fashion (see Appendix B for another interpretation). In addition the Variable Attention Model views capacity as a resource (attention, stimulus quality, etc.) that varies across trials, but remains fixed within a trial. A summary of the models and their underlying assumptions is found in Table 3.

INSERT TABLE 3 ABOUT HERE

Below we develop the Fixed Path Poisson Model, and then demonstrate how it can account for several conditional probability statistics that indicate the existence of positive inter-item dependencies. Previous data as well as the Loftus, et al. (1993) data, contain non-uniform serial position effects. The Fixed Path Poisson Model predicts strong serial position effects that qualitatively are in line with the data. As it turns out those predictions are too extreme for the data, and we will alter this model to accommodate these effects. However, this new model contains the same dependency structure as the Fixed Path Poisson Model, the simplicity of the latter recommends it for first exposition.

The Fixed Path Poisson Model

Conceptually the Fixed Path Poisson Model is quite simple: upon stimulus onset, an observer begins viewing the first digit. The time to acquire enough information to correctly report a digit is exponentially distributed. Once the first digit is acquired, processing begins on the second digit, and continues through to the fourth digit. Processing ends upon stimulus offset or the completion of the fourth digit. If necessary, the observer then guesses the identity of any digits that have not yet been perceived. The rate of processing of an individual item is designated by the parameter λ . Note that in the standard serial interpretation of this model, processing of a single item is unaffected by the number of other items in the display. That is, λ is unaffected by other items in the display, and thus this model is identified in Figure 1 as Unlimited Capacity. However, in the parallel interpretation, the processing rate is allocated to all of the items in the display, and thus the initial processing rate of each item (assuming equal allocation) is λ/n . As a result, the processing rate of an individual item does depend on the total number of items, and we designate this model as Limited Capacity in Figure 1. In the discussion below, we will stick with the standard serial interpretation; a discussion of the parallel mimicking form is found in Appendix B.

The observer always moves from left to right, which identifies this Poisson Model as the Fixed Path Poisson Model. For instance, in some studies (e.g., Shibuya & Bundesen 1988) performance declines from left-to-right for horizontal linear arrays, although with larger arrays other factors such as lateral interference may introduce non-monotonicities (see Townsend, 1981). This type of result suggests the possibility of a serial mechanism with a preferred (perhaps fixed) order of processing, as instantiated in the Fixed Path Poisson Model. The strict left-to-right processing assumption is, of course, very strong, and later, by looking at the probability of a correct item beyond the first error, we will disconfirm the Fixed

Path Poisson Model. In a subsequent section we develop a version of the Poisson model that allows the observer to take different paths through the four digits. Because the extended model relies on the same basic dependency structure as the Fixed Path Poisson Model, we develop this model below with the understanding that it will be extended to relax the strict left-to-right processing assumption.

Before detailing the mathematics of the Fixed Path Poisson Model, we should point out that this model is easy to test in a qualitative fashion. Under the strict left-to-right processing assumption, once an error is made, digits to the right of the error can only be obtained by guessing. Therefore, once an error is made, the probability of obtaining a correct digit cannot exceed the guessing rate. More formally, this relation must hold:

$$P(C_j | \overline{C}_i) = 0.1 \quad \text{Eq. 3}$$

where $j > i$, and \overline{C}_i indicates that the digit in location i was incorrect. As shown in Figure 3, this prediction is violated for all three observers. The Independence model predicts that the three curves should be equal and increase at a rate equal to the marginal performance curves shown in Figure 2. The Figure 3 data clearly disconfirm the Independent Sampling Model as well. To account for this statistic and other aspects of the data, we will amend the Fixed Path Poisson Model to include a mechanism by which subjects can take any one of the 24 processing paths through the 4 stimulus locations. This new model, which we refer to as the Weighted Path Poisson Model, will be a direct extension of the Fixed Path version.

Townsend (1981) has derived predictions for the Fixed Path Poisson Model. The probability that a digit in the i th position is reported correctly is

$$P(C_i) = \sum_{j=0}^{i-1} P(j, t, t_o) \frac{1}{10} + \sum_{k=i}^{\infty} P(k, t, t_o), \quad \text{Eq. 4}$$

where

$$P(j, t, t_o) = \frac{[\lambda(t - t_o)]^j e^{-\lambda(t - t_o)}}{j!}, \quad \text{Eq. 5}$$

is the Poisson probability that j letters have been completed by time t . As stated above, the parameter λ is the poisson rate parameter. The parameter t_o is a processing delay parameter, similar to the Liftoff parameter of Eq. 1. The first sum in Eq. 4 is multiplied by $1/10$ to account for unperceived digits that are

correctly guessed⁴. This model has two free parameters, λ and t_0 , and when they are set to the best-fitting values shown in Table 1, they generate the short-dashed curves in Figure 2. The root-mean-squared-error (RMSE) was computed for the Independent Sampling Model and the Fixed Path Poisson Model to evaluate the goodness of each fit. Durations below the processing-delay parameter were not included in the evaluation of the models.

INSERT FIGURE 3 ABOUT HERE

Each of the Independent Sampling Model fits in Figure 2 has a slightly larger RMSE than the associated Fixed Path Poisson Model fit. Based on the analysis of the marginal probabilities along the Fixed Path Poisson Model gives a somewhat better account of the data. However, the fact that two quite different models should give such similar predictions gives credence to our suggestion that analysis of the marginal probability predictions may not always discriminate well between candidate models.

The Weighted Path Poisson Model

The Fixed Path Poisson Model makes the strong assumption that observers invariably begin working on the left-most position and work to the right. This path represents only one of $4!=24$ possible paths that the observer might take, and evaluating each of these paths, or combinations of paths, would require more parameters than data points. An alternative is a model that computes the probability of each of these 24 paths from just 4 parameters that express the likelihood that a given stimulus location appears in different processing positions. We can take a cue from Bundesen (e.g., 1993) in which path selection is related to Luce's Choice Model (e.g., 1959), although in Bundesen's model, processing is assumed to be parallel and independent⁵. Even our parallel model that is equivalent to this serial interpretation differs from Bundesen's parallel model, in that reallocation of capacity produces a positive dependence in our model (see Appendices A and B). This new model, termed the Weighted Path Poisson Model, computes the marginal probability of correctly identifying the digit in stimulus location i as,

⁴One difference between the paradigms used by Loftus et. al. and by Townsend was the generation of the stimuli. Townsend's stimuli could not contain repeats within the five digits he displayed, which lead to more complicated guessing terms, and introduced a mild negative dependency into the Bounded Performance Model. The Loftus stimuli were generated allowing replacements, with the exception of 3 or 4 of the same digit in a row. This eliminates less than 2% of the potential stimuli, making the stimulus generation process a good approximation to sampling-with-replacement.

⁵ Bundesen (1993) further investigates a generalization of independent exponential models where each item's processing is governed by distributions hazard functions that were proportional to one another.

$$P(C_i) = \sum_{n=1}^4 u_{i,n} \left[1 - \sum_{j=0}^{n-1} P(j,t,t_0) + \left(\sum_{j=0}^{n-1} P(j,t,t_0) \right) g \right] \tag{Eq 6}$$

were $P(j, t, t_0)$ comes from Eq 5 and $u_{i,n}$ represents the probability that a given stimulus location i appears in processing position n of the Poisson process. The rest of the term represents the probability that the Poisson process reaches position n , or if not, the digit is obtained through the guessing process with probability g . The value of $u_{i,n}$ is determined by four weight values $w_i, i = 1, 2, 3, 4$, which are positive and constrained to sum to one. The probability that stimulus location i is processed in position n of the Poisson process is,

$$u_{i,n} = \begin{cases} w_i & n = 1 \\ \sum_{\substack{j=1 \\ j \neq i}}^4 w_j \frac{w_i}{1 - w_j} & n = 2 \\ \sum_{\substack{j=1 \\ j \neq i}}^4 \sum_{\substack{k=1 \\ k \neq i, j}}^4 w_j \frac{w_k}{1 - w_j} \frac{w_i}{1 - w_j - w_k} & n = 3 \\ \sum_{\substack{j=1 \\ j \neq i}}^4 \sum_{\substack{k=1 \\ k \neq i, j}}^4 \sum_{\substack{l=1 \\ l \neq i, j, k}}^4 w_j \frac{w_k}{1 - w_j} \frac{w_l}{1 - w_j - w_k} & n = 4 \end{cases} \tag{Eq 7}$$

An example illustrates this computation. Suppose that weights $w_1 \dots w_4$ are .45, .30, .20 and 0.05. Given these weights, stimulus location 2 is processed first ($n=1$) 30% of the time. This leaves three stimulus locations to be processed. The probability that stimulus location 1 is processed second ($n=2$) given location 2 was processed first is $.45 / (.45 + .20 + .05)$. The probability that location 3 is processed third ($n=3$) is $.20 / (.20 + .05)$. This leaves only location 4 to be processed last, which it is with probability $.05 / .05$, or 1.0. Of course for short stimulus durations, the Poisson process may not get all the way to the last position, and if not, the observer guesses the remaining digits.

The right panels of Figure 3 show the fits of the Weighted Path Poisson Model to the statistic $P(C_i | \text{previous error})$. These fits were generated via a simulation technique that is described in a subsequent section, but are presented here to demonstrate the adequacy of the Weighted Path Poisson Model when fitting correct items beyond an error.

Conditional Probability Analyses

Having developed the Weighted Path Poisson Model and shown how it can account for one kind of dependency (correct items beyond an error), we now extend it to account for additional dependency statistics. Our evaluation of candidate models takes a two-pronged approach. We will consider two statistics that provide what we feel are intuitive insights into the structure of the dependencies, since the log-likelihood analyses do not provide the direction of the violations of independence. The statistics that we develop first do not allow the powerful hypothesis testing that log-likelihood analysis provide. Therefore we will develop the conditional probability analyses and then perform log-likelihood analysis to confirm the results of the conditional statistics. Nevertheless, we expected that any model that failed our coarser statistics would not perform well on the finer grained tests.

Our conditional probability statistics consider the probability that a digit is reported correctly, given that one or more other digits in the same trial are reported correctly. A positive dependency implies that knowing that a digit was correct increases the likelihood that another digit was also correct. A negative dependency implies that knowing that a digit was correct decreases the likelihood that another digit was also correct. Townsend (1981, Townsend & Ashby, 1983) has proposed two statistics that quantify the degree and direction of inter-item dependency present. These statistics do not require explicit parameter estimation. They compute two kinds of conditional probabilities that proved useful in the Townsend (1981) study for testing three principled models predicting qualitatively and quantitatively different dependencies. These summary statistics can provide an intuitive overview of the nature of the dependencies and we have found them helpful in model development.

The first statistic considers the probability that one item is acquired given that another is perceptually acquired, which is $P(C_i|C_j)$. The size of $P(C_i|C_j)$ depends upon stimulus duration. For example, a stimulus duration that produces chance or asymptotic performance cannot produce inter-item dependencies, because both the marginal and conditional probabilities will be 1.0. However, conditions that produce marginal performance near .5 allow for the possibility of large inter-item dependencies. One way to remove the effects of the level of marginal performance is to subtract it from $P(C_i|C_j)$. Thus $Ave[P(C_i|C_j)-P(C_i)]$, $i \neq j$, is computed and averaged over i and j for each separate exposure duration. We plot $Ave[P(C_i|C_j)-P(C_i)]$ as a function of $P(C_i)$. Both of these statistics are functions of their parameters and time in such a way that when time increases from very small to very large, the average difference

sweeps out its magnitude characteristic and the overall curve does not depend on any particular parameter value (see Townsend, 1981, for quantitative detail). In the Poisson models the rate of item processing is $\lambda(t-to)$. Hence, regardless of the value of λ , the same curves for these probability functions result as t varies from very small to very large. A comparable statement holds for the other models under consideration, indeed for all the models that Townsend (1981) analyzed as well, although obviously the exact form of the functions differs across models.

Conditional probability derivations for the Fixed Path Poisson Model are found in Townsend & Ashby (1983). The Independent Sampling Model is straightforward: the defining characteristic of this model is that items in the display are acquired independently. Thus the Independent Sampling Model predicts $\text{Ave}[P(C_i|C_j)-P(C_i)]$ to be zero for all values of i and j in the present study, and non-zero values of this statistic tend to disconfirm the model. Model predictions and dependencies from the three participants in Experiment 1 of Loftus, Busey & Senders (1993) are plotted in Figure 4. Each observer contributes one point to this graph for every stimulus duration that produces above-chance performance. Contrary to the predictions of the Independent Sampling Model, the data show clear positive dependencies for durations that produced above-chance performance. A sign test on the 16 points from the three observers reveals that 12 are above zero, which is significant at $\alpha = 0.05$. These dependencies clearly reject the Independent Sampling Model. This finding is in contrast to the mild negative dependencies reported by Townsend (1981) for a similar task. However, the mild negative dependencies in the earlier study were entirely accounted for by guessing influences: because the digits were selected without replacement, knowing that a digit was presented increased the likelihood of guessing subsequent digits, because certain numbers could be eliminated. When this "selection without replacement" guessing structure was included in the Bounded Performance Model, the predicted inter-item dependencies corresponded to those in the data (see Footnote 2). Individual values and predictions from the Poisson models are found in Table 4.

INSERT FIGURE 4 ABOUT HERE

INSERT TABLE 4 ABOUT HERE

A second inter-item dependency statistic is $P(C_i|k)$, the probability of getting item C_i correct given that exactly k other items were correct on that trial, where k ranges from 0 to 3 and i ranges from 1 to 4. These probabilities are averaged over the four serial positions (abbreviated as $P(C|k)$), and plotted against k . Separate curves are computed for each stimulus duration. These data are shown in Figure 5. The light

curves represent the obtained performance, the dark curves are predictions from the Independent Sampling and Fixed-Path Poisson models. These theoretical predictions were generated using the best-fitting parameters from Figure 2 for each model and observer. Because the parameters from the models were fit to the marginal data, these inter-item dependency data predictions have no remaining free parameters. Data and theoretical predictions are omitted where little data contribute to that point.

INSERT FIGURE 5 ABOUT HERE

The Fixed Path Poisson Model provides a much better fit to the data than the Independent Sampling Model. The Figure 5 data (solid curves) indicate the presence of some form of positive inter-item dependencies, which is again contrary to the little or no inter-item dependencies reported by Townsend (1981) beyond those accounted for by guessing. These dependencies clearly reject the Independent Sampling Model. The Fixed Path Poisson Model predictions come quite close to the overall pattern of data, for all three participants.

Details of the dependency structure for the Weighted Path Poisson Model will be given shortly. However, we describe this model's predictions for the present statistics for comparison sake.

Weighted Path Poisson Model Dependency Predictions

Predictions for the dependency statistics $Ave[P(C_i|C_j)-P(C_i)]$, $P(C|k)$ and $P(C_i|previous\ error)$ were generated by performing simulations using the estimated column weights w_1, \dots, w_l , the processing rate λ and the pre-processing delay t_0 . Each condition was simulated for 15,000 trials using the same digit set as the original experiment. This simulation involves picking a path for the Poisson model to take on a particular trial, and then applying the Fixed-Path Poisson Model to this particular path according to Eqs 2 and 3. The probability of any given path $[J_i]$ can be directly computed from Eq 10 below. Once a path is chosen, correct and incorrect digits are assigned according to Eqs 2 and 3 and the weight, rate and delay parameters for each observer⁶ (see Table 1). As a test of the simulation procedures, the analysis produced a data set that contained four-way joint probabilities which were quite close to those derived analytically from the Weighted Path Poisson Model using the appropriate parameters.

⁶It is easy to show that this assumption of selecting the processing path is equivalent to one where each successive item position is picked at each stage of processing, en route, as it were.

Figure 6 shows the predictions for the Weighted Path Poisson Model for the three participants for the $P(C|k)$ statistic. The RMSE's are all similar to those of the Fixed-Path Poisson Model, as expected, and all are lower than those of the Independent Sampling Model. In addition, Table 4 lists the predictions for the statistic $Ave[P(C_i|C_j)-P(C_i)]$ for the Weighted Path Poisson Model, which are all similar to the predictions made by the Fixed Path Poisson Model, also as expected. The right panels of Figure 3 shows the predictions for the $P(C_i|previous\ error)$ statistic.

INSERT FIGURE 6 ABOUT HERE

Log-Likelihood Tests of the Independence and Weighted Path Models

The two conditional probability statistics described above demonstrate that the data contain positive dependencies that are consistent with the Fixed-Path Poisson Model. These statistics demonstrate the direction of the dependencies and by inspection produce good fits to the empirical data. More rigorous model testing is provided by log-likelihood tests performed on the four-way joint probabilities. These joint probabilities exhaust all inter-item dependency information and, assuming inter-trial independence, form a set of sufficient (complete) statistics for the data set and for testing dependencies in particular. The strongest test of the Independence and Poisson models comes from the model's ability to predict the four-way joint random variables $[X_i] = (X_1, X_2, X_3, X_4) = (0000, 1000, 0100, \dots, 1111)$, where X_i is the random variable at location i where a '1' signifies correct and '0' signifies incorrect. Four way joint probabilities can be computed from the response frequencies by dividing by the total number of trials at a particular stimulus duration.

Table 5 contains the counts of the joint events for the three observers for the 8 stimulus durations and 16 response types.

INSERT TABLE 5 ABOUT HERE

Four-way joint probability predictions for the Weighted Path Poisson Model follow directly from the derivations from the Fixed Path Poisson Model, and therefore we develop the predictions from the Fixed Path Model first. Under the fixed-path assumption, an item can only be obtained to the right of an error through the guessing process. For instance,

$$P(X_1 = 1 \cap X_2 = 0 \cap X_3 = 1 \cap X_4 = 0) = P(0,t,t_o)(g)^2(1-g)^2 + P(1,t,t_o)(g)(1-g)^2 \tag{Eq 8}$$

where $P(j,t,t_o)$ is the probability of getting j items by the Poisson process at time t from Eq. 5, g is the guessing rate, and t_o is the pre-processing delay parameter. The entire expression in Eq 8 computes the probability of getting none of the digits by the Poisson process but guessing two correctly or getting the first digit by the Poisson process and correctly guessing the third digit. These two events are mutually exclusive and therefore sum to produce the overall joint probability. We cannot get both digits by the Poisson process (which would include the term $P(2,t,t_o)$) because once an error is made, digits to the right may only be obtained by guessing.

A straightforward way to produce predictions in general for the Fixed Path Poisson Model is to designate the position just before an error, if any occurs, as r . Then letting s be the number of correct responses after r (s may equal 0) we can write any such sequence as

$$P_{FPP}(r + s \text{ correct and } 4 - (r + s) \text{ incorrect in exact order}) = \sum_{j=0}^r P(j,t,t_o) g^{r+s-j} (1-g)^{4-r-s}, r+s < 4 \quad \text{Eq 9}$$

where we subscript this probability by FPP to designate this as the joint probability derived from the Fixed Path Poisson Model. Note that if $r=4, s=0$, we define

$$P_{FPP}(X_1 = 1 \cap X_2 = 1 \cap X_3 = 1 \cap X_4 = 1) = \sum_{j=0}^3 P(j,t,t_o) g^{4-j} + \left(1 - \sum_{i=0}^3 P(i,t,t_o) \right)$$

in keeping with the truncated Poisson distribution that results from the fact that the observer cannot get more than four digits correct.

Predictions for the four way joint probabilities can be obtained for the Weighted Path Poisson Model by computing the four way joint probabilities for the Fixed Path Poisson Model and then using these to find the correct probability for any processing path permutation. As with the Fixed Path Poisson Model, define $\mathbf{X}_i = \{0,1\}$, $i=1..4$, which represents whether the digit was incorrect or correct on *stimulus location* i . Define $\mathbf{Y}_i = \{0,1\}$, $i=1..4$, which represents whether the digit was incorrect or correct on *processing position* i in the Poisson process.

The four-way joint probability computations require computing probability of any one of the 24 possible paths. Define $[\mathbf{J}_i]$, $i=1..4$ as one of the possible paths. The \mathbf{J}_i indexes the stimulus location

corresponding to processing position i . The probability φ of path $[J_i]$ can be directly computed from the weight values:

$$\varphi([J_i]) = w_{J_1} \frac{w_{J_2}}{(1 - w_{J_1})} \frac{w_{J_3}}{(1 - w_{J_1} - w_{J_2})} \quad \text{Eq 10}$$

where $w_a..w_d$ are the weight values $w_1..w_4$ from Eq 9 above. In the previous example illustrating Eq 9 above, $[J_i] = (2, 1, 3, 4)$ and $\varphi([J_i]) = w_2 \frac{w_1}{(1 - w_2)} \frac{w_3}{(1 - w_2 - w_1)} = .3 \frac{.45}{(1-.3)} \frac{.2}{(1-.3-.45)} = .154$.

The four way joint probabilities for the Weighted Path Poisson Model are then defined as

$$P_{WPP}(X_1 \cap X_2 \cap X_3 \cap X_4) = \sum_{\{J_i\} = (J_1, J_2, J_3, J_4) \in \Pi} \varphi([J_i]) P_{FPP}(Y_1 = X_{J_1} \cap Y_2 = X_{J_2} \cap Y_3 = X_{J_3} \cap Y_4 = X_{J_4} | [J_i]) \quad \text{Eq 11}$$

where Π represents the set of all 24 permutations, $\varphi([J_i])$ computes the probability of path $[J_i]$ from Eq 10 and P_{FPP} gives the appropriate four way joint probability from the Fixed-Path Poisson Model from Eq 6. The J_i subscripts provide a redirection of the Poisson process so that instead of moving strictly from left to right, the process moves through any one of the 24 possible orders, while still preserving its serial nature of processing one stimulus location at a time. For example, to compute the probability of

$P_{WPP}(X_1 = 0 \cap X_2 = 1 \cap X_3 = 1 \cap X_4 = 0)$, first consider the case where the third stimulus location is processed first, followed by the second, first and fourth locations. Define $\mathbf{X}_i = (0, 1, 1, 0)$ and $\mathbf{J} = (3, 2, 1, 4)$ as noted, and then find the probability

$P_{WPP}(X_1 = 0 \cap X_2 = 1 \cap X_3 = 1 \cap X_4 = 0 | [J_i] = (3, 2, 1, 4)) = P_{FPP}(Y_1 = 1, Y_2 = 1, Y_3 = 1, Y_4 = 0)$ from the Fixed Path Poisson Model. Finally, weight this by the probability of taking processing path 3, 2, 1, 4 via Eq 10 and the 4 weight values. This process is repeated for all 24 possible processing paths and the results are summed to compute the four way joint probability $P_{WPP}(X_1 \cap X_2 \cap X_3 \cap X_4)$.

The Weighted Path Poisson Model has 5 free parameters (t_0 , λ , and w_1, w_j, w_k). The parameter w_1 is not included in the estimation since the weights sum to 1.0.

A fair comparison between the Weighted Path Poisson Model and the Independent Sampling Model allows different processing rates on the four stimulus positions for the Independent Sampling Model. As with the Fixed-Path Poisson Model, define $\mathbf{X}_i = \{0,1\}$, $i=1..4$, depending upon whether the digit in

stimulus location i through l is incorrect or correct. For the Independent Sampling Model, the joint probability for a given \mathbf{X}_i is defined to be

$$P_{Ind}(\mathbf{X}_1 \cap \mathbf{X}_2 \cap \mathbf{X}_3 \cap \mathbf{X}_4) = \prod_{j=1}^4 (P(C_j)^{X_j} (1 - P(C_j))^{1-X_j}) \quad \text{Eq 12}$$

where $P(C_j)$ is the probability of obtaining a digit in location j from Eq 1, using either the random sampling process with sampling rate $1/c_j$ or by guessing. For example,

$$P_{Ind}(\mathbf{X}_1 = 1 \cap \mathbf{X}_2 = 1 \cap \mathbf{X}_3 = 0 \cap \mathbf{X}_4 = 0) = P(C_1)P(C_2)(1 - P(C_3))(1 - P(C_4)) \quad \text{Eq 13}$$

where $P(C_j)$ is the probability of obtaining a digit in location i from the independent sampling process with rate parameter $1/c_j$ from Eq 1 or by guessing. The four-way joint probability is the product of these four probabilities, under the independence assumption. Using Eq 13 we computed the four-way joint probabilities for the independence model assuming 5 free parameters: t_0 and 4 processing rates $1/c_1, 1/c_2, 1/c_3, 1/c_4$.

Having derived predictions for the four-way joint probabilities for the two models, we can now test these by maximizing the log-likelihood statistic via parameter estimation techniques. The log-likelihood statistic is computed as

$$\log \left[\prod_{dur=1}^{tdurs} \prod_{X_1=0}^1 \prod_{X_2=0}^1 \prod_{X_3=0}^1 \prod_{X_4=0}^1 DataJnt(dur, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)^{N(dur, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)} \right] - \log \left[\prod_{dur=1}^{tdurs} \prod_{X_1=0}^1 \prod_{X_2=0}^1 \prod_{X_3=0}^1 \prod_{X_4=0}^1 TheoryJnt(dur, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)^{N(dur, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)} \right] \quad \text{Eq 14}$$

where $N(dur, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)$ is the number of occurrences in the data for a particular four way joint event, stimulus duration and observer, as given in Table 5. $DataJnt(dur, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)$ is the relative frequency of this joint event occurring in the data for a particular stimulus duration dur and observer, which is computed by dividing each joint count by the total number of trials at that stimulus duration. $TheoryJnt(dur, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)$ is the predicted four-way joint probability for a given stimulus duration and particular configuration of correct and incorrect positions $\mathbf{X}_1 \dots \mathbf{X}_4$ for the Weighted Path Poisson or Independence models. When multiplied by 2.0, this statistic is distributed as chi-square. The number of degrees of freedom equals $(16-1)$ times the number of above-chance stimulus durations, minus the number of free parameters (5 for both models). The value of $tdurs$ is derived from the number of stimulus

durations that produce above-chance and below-ceiling performance, since the inter-item dependencies do not discriminate between models at these extremes.

Predictions for the two models were derived by computing the four-way joint probabilities by either Eq 11 or 12 and minimizing Eq 14. Minimization was performed in Mathematica using the FindMinimum procedure. Table 6 contains the maximum likelihood values obtained for the two models for the three observers. As anticipated from the original tests of independence, the Independent Sampling Model is rejected for all three observers. The Weighted Path Poisson Model does better than the Independent Sampling Model for all 3 observers, and substantially so for Observers SS and EF. In addition, the model is not rejected for Observer EF. The log-likelihood statistic is fairly close to the critical χ^2 value for the other two observers.

INSERT TABLE 6 ABOUT HERE

Variable Attention and Independent Sampling

The Weighted Path Poisson Model clearly handles our two dependency statistics and the four-way joint probabilities better than the modal independence class of models, and comes much closer to predicting the serial position effects than the Fixed Path Poisson Model. It is obviously not feasible to anticipate all other models, based for example, on distributions other than the Poisson, that may yield positive dependencies. However, it seemed important to rule out, if possible, a model that could predict positive dependencies by artifact, in the sense that its across-trial, behavior might produce dependencies, even if the within-trial stochastic process were independent.

In this model, attention (or some other facility positively related to performance) is assumed to vary across trials, but remains constant within a trial. Such variable attention models have attracted the notice of researchers for many years (e.g., Atkinson, 1963; Norman, 1964), although there is some evidence against this notion, at least in featural dependencies (e.g., Townsend, Hu and Kadlec, 1988). Van Zandt and Ratcliff(1995) have recently argued for a similar concept within the context of diffusion processing models. The result is, of course, a probability mixture over trials of the predicted function of the varying parameter. It does seem a natural idea that people's attention might wax and wane over a prolonged testing session. Alternatively, such variations in processing may come from the stimulus itself, or low-level mechanisms such as the position of fixation during a trial. Whatever the mechanism, the result is that the processing of all four stimulus locations can be either higher than average or lower than average on a given

trial. This effectively correlates the processing of the individual items, even if processing occurs independently for each item within a given trial.

One way to incorporate the idea of variable attention across trials is to append this mechanism to the Independent Sampling Model. Currently, the ISM has 4 processing rates on the four stimulus locations. If the observer is currently in a high attentional state, we would expect that processing would increase at all four locations, but proportionately more at locations that were proceeding at a faster rate. Thus attention can be thought of as a multiplicative weight that adjusts the individual processing rates on the four stimulus locations to reflect the current state of attention. One could add a variable attention component to any model (and below we add this component to the Weighted Path Poisson Model as well). However, for simplicity, we will refer to the variable attention version of the Independent Sampling Model as the Variable Attention Model. While this model still includes an independent sampling mechanism within a trial, it now predicts positive inter-item dependencies.

First, we test a model approximating the case where the attentional states constitute a continuous random variable that is approximately normally distributed. For computational ease, we chose a 5-state distribution that is approximately normally distributed. This captures the characteristics that attention remains at a modal value for many of the trials, but can fluctuate to higher or lower attention states on other trials. More states were not used because for each state the model predictions must be re-calculated and this became computationally intensive. We also consider a simpler case below in which attention is only in one of two states. Within the 5-state model, the mean of the attentional distribution would reflect the processing rate for the average attentional state. To develop this model, we approximated a normal distribution with a binomial distribution with $n=4$. This provides an approximately bell-shaped density function with 5 discrete states. To compute overall performance, we simply compute performance for the four-way joint probability when in each of 5 attentional states and perform a weighted average of the joint probabilities.

The mean of the binomial function, μ , was a free parameter that typically ended up at a value near 0.5. Thus for each of the 5 attentional states \mathbf{a} , the probability of being in a particular state $\alpha(\mathbf{a})$, $\mathbf{a}=\{0..4\}$ is given by:

$$\alpha(\mathbf{a}) = \binom{4}{\mathbf{a}} \mu^{\mathbf{a}} (1 - \mu)^{4 - \mathbf{a}} \quad \text{Eq 15}$$

The spread of attention, or what might be thought of as the variance of attention across these 5 attentional states, is given by a free parameter σ . The combination of the attentional weight scaling the processing rate determines what we term the *pace* of processing. For stimulus location i , the pace of information acquisition is,

$$pace_i = (1 - (\mathbf{a} - 2)\sigma)\lambda_i \quad \text{Eq 16}$$

When $\mathbf{a}=2$, we multiply each rate by 1.0, and thus the processing pace occurs at the average rate assigned to each stimulus location. When $\mathbf{a}=1$, we reduce the attention multiplier by one standard deviation unit σ , such that each rate is multiplied by $(1.0 - \sigma)$. Once the rate for each location is determined, processing is assumed to occur in parallel and independently, and the four-way joint probabilities are given by Eq 12. For all five attentional states, the overall joint probability for a given \mathbf{X}_i is determined by,

$$P_{VAM}(\mathbf{X}_1 \cap \mathbf{X}_2 \cap \mathbf{X}_3 \cap \mathbf{X}_4 | \mathbf{a}) = \sum_{\mathbf{a}=0}^4 \alpha(\mathbf{a}) \left(\prod_{j=1}^4 \left(P(C_j | \mathbf{a})^{X_j} \right) \left(1 - P(C_j | \mathbf{a})^{1 - X_j} \right) \right) \quad \text{Eq 17}$$

where to compute $P(C_j | \mathbf{a})$, we scale the rate parameter by the appropriate attentional state weighting parameter,

$$P(C_j | \mathbf{a}) = (1.0 - e^{-(d-L)(1/c - (1 - (\mathbf{a} - 2)\sigma))}) + g(e^{-(d-L)(1/c)(1 - (\mathbf{a} - 2)\sigma)}) \quad \text{Eq 18}$$

The end result is a distribution of attentional weights that scale the processing rates. For a given trial, the distribution of attention weights is sampled, and the chosen value is used to scale the four processing rates on the four stimulus locations. This produces four processing paces for the four locations. Note this is equivalent to assuming that once an attentional state is "chosen" for a trial, the associated attention weight is used to scale all four processing rates.

This model was fit to the four-way joint probability data, and produces surprisingly (at least to us) good results, as shown in Table 6. Even allowing for the addition of two free parameters, the Variable Attention Model accounts for the dependency data even better than the Weighted-Path Poisson Model, and in all three cases the log-likelihood value is below the critical χ^2 value. Thus on the basis of the model fits, we must conclude that the Variable Attention Model is the superior model. Nevertheless, it should be

observed that these good fits require extremely large shifts in attention. All three observers have σ values greater than .45; .5 is the maximum value that σ could take, since values greater than .5 imply negative processing pace.

When evaluating the fit of any model, it is important to ask whether the parameter values that enable good fits are reasonable. For example, consider Observer EF's performance at the 80 ms presentation. The model assumes that she is in one of 5 attentional states with the following probabilities (from the lowest to highest attention states): {0.035, 0.18, 0.36, 0.32, 0.10}. The probability that she obtains a correct digit (averaged across all 4 locations) when in these 5 states is {0.24, 0.28, 0.35, 0.52, 1.0}. This is equivalent to changing the stimulus duration from 70 ms to 200 ms! Such extreme shifts in attention across trials seem improbable to us in the conventional sense of attention waxing and waning within a block of trials. This aspect will be discussed further below.

Fluctuating attention would seem likely to vary slowly over trials. In order to evaluate this possibility, we performed an autocorrelation on the trial-by-trial data by first subtracting off the mean performance level for each condition on each trial, and then autocorrelating the data at lags up to half of the experiment. The results are shown in Figure 7. None of the three observers showed any structure at all in the autocorrelation function; the only deviations from the confidence band are those expected due to chance. This was much the same result as in Townsend et al. (1988), and suggests that a slowly fluctuating attentional system does not underlie the positive inter-item dependencies observed in our data.

INSERT FIGURE 7 ABOUT HERE

It seems plausible that attention may fluctuate in a more binary, rather than continuous fashion. In order to check the idea of a binary rather than finer-grained fluctuation of attention, we developed a two-state attention model in which the observer was assumed to be either in a low or a high attentional state with some probability. This also produced worse fits than the Variable Attention Model described above, and also required wide shifts in attention across trials in order to account for the dependencies. Even so, this model did perform better than the Weighted Path Poisson Model. We also appended a variable attention mechanism to the Weighted Path Poisson Model, but found that it did not markedly improve the fits. We also extended the Variable Attention Model developed in Eqs 15 and 16 to include 41 individual states, binomially distributed. This provides a much better approximation to a normal distribution. This

model produces results that are equivalent to the 5-state version, and demonstrates that our choice of the number of individual states does not affect the Variable Attention Model's fit to the dependency data.

INSERT TABLE 7 ABOUT HERE

The maximum-likelihood values for the Variable Attention Model fits demonstrate that we cannot reject this model, although the means by which it derives its good predictions seem somewhat suspect, at least under the usual interpretation of attention varying gradually across trials. However, as intimated earlier, there might be other influences that could change across trials but remain fixed within a trial. For example, although all three subjects were well-practiced, the position of the eyes during a trial may have influenced whether performance was good or poor overall on this trial. Other sources of variability may come from the stimulus itself, such that certain number combinations might be easier to read than others. Also, as noted earlier, the perceptual integrity of the incoming information might vary. If so, it might well be independent from trial to trial (and thus produce zero or irregular autocorrelation functions) and perhaps even be normal in distribution. Perhaps some combination of these various factors could be responsible for high and independent variability across trials.

Although we have contrasted the serial Weighted Path Poisson Model with the parallel Independent Sampling Model and its variable attention variant, we wish to reemphasize that both models can be mimicked by models of other architectures. As noted by Hughes & Townsend (1998), accuracy alone may give little information about process architecture. The focus of the present discussion is on the ability of the various models to account for inter-item dependencies, which is somewhat independent of the architecture of the model. A limited discussion of the ability of the present serial model and parallel architectures to mimic one another can be found in Appendix B.

Discussion and Conclusions

We conclude, based on log-likelihood analyses of the four-way joint probabilities as well as analyses of the two conditional-probability statistics $\text{Ave}[P(C_i|C_j)-P(C_j)]$ and $P(C|k)$, that the Independent Sampling Model, or in fact any of its modal model cousins, cannot account for several aspects of performance in the digit-recall task. In particular, a non-parametric independence statistic verifies the existence of positive inter-item dependencies that are contrary to the Independent Sampling Model. The Fixed-Path Poisson Model could account for the two inter-item dependencies but not the serial position statistics and the probability of obtaining a digit to the right of an error.

To account for these last two statistics we developed the Weighted Path Poisson Model. This model computes the probability of taking one of 24 possible paths through the 4 digit locations based on three weights that determine the likelihood that a given digit location appears in different processing positions. We tested the Weighted Path Poisson Model against the Independent Sampling Model using log likelihood statistics and found that the Weighted Path Poisson Model performed better for all three observers. The Independent Sampling Model was rejected for all three observers, while the Weighted Path Poisson Model produced a log-likelihood value that was below the critical χ^2 value for one observer and near the critical value for the other two observers.

Finally, we developed a version of the Independent Sampling Model that assumes that inter-item dependencies result from attention varying across trials, termed the Variable Attention Model. This model assumes that processing is independent across channels within a trial, but that due to variation in attention across trials, an overall positive dependence is produced. This model also accurately captures the obtained serial position effects. Overall this model did an even better job than the Weighted Path Poisson Model, and actually produced log-likelihood values that were below the critical values. However, it derives its good fits by permitting extremely large shifts in attention that seem too massive to be produced by the traditional assumption of waxing and waning attention across trials. In addition, no evidence for systematic attentional shifts across trials were found in the autocorrelation plots. As a result, this model seems difficult to accept without assuming that other mechanisms are responsible for across-trial variability, mechanisms that do not produce correlated shifts in attention. A reasonable interpretation of these attentional shifts would come from a situation in which variations in signal quality, perhaps in conjunction with attentional oscillation, might be able to accommodate such large swings in performance combined with the null autocorrelation results.

One possible explanation for the rejection of the Weighted Path Poisson Model for Observers TB and SS may be that the positive dependencies were not quite as strong as those predicted by that model. Inspection of Table 2 and Figure 5 reveals that although the dependencies are clearly positive, they are not quite as large as the model's predictions. One way to account for these effects would be to include a limited-capacity short-term memory store perhaps comparable to the fixed sample size buffer notion of Townsend (1981) or that proposed by Bundesen (1990; see also Shibuya & Bundesen, 1988), at the back end of the Poisson process. Limiting the processing capacity produces negative dependencies, which

might offset the strong positive dependencies produced by the Weighted Path Poisson Model⁷. Obviously such notions are post-hoc at this point.

The observed dependencies are positive, contrary to those reported by Townsend(1981), the only other full report study to analyze dependencies so far. Townsend's data showed slight negative dependencies, attributable to guessing, that are consistent with an independent sampling model. At this point, it not known what led to the differences in findings in our present analysis verses those in the Townsend (1981) study. Although the stimuli in both experiments were post-masked, the Loftus et al. (1993) stimuli were very low contrast (around 5%), while the Townsend (1981) stimuli were black typewritten letters on white cards presented at high luminances. Also, pilot data from that study indicated that presentation of one or two stimuli in any of the display locations led to virtually perfect performance. This may have placed the Loftus et al. (1993) stimuli in a data-limited domain, while the bright stimuli of the Townsend (1981) study may have been in a process-limited domain due to time constraints. In addition, the Townsend stimuli were much more widely separated and the letters subtended a smaller visual angle than the Loftus, et al (1993) stimuli, possibly encouraging perceptual independence. With regard to the data limited hypothesis, the present dependency results are more compatible with those of Townsend and colleagues (e.g. Townsend, Hu & Evans, 1984) which found uniformly strong positive dependencies, albeit among full report of features in a pattern recognition, rather than a full report of letters, experiment.

Although the Poisson class of models, whether given a serial or parallel interpretation, clearly accommodated the Loftus et al data better than the modal independence type of model, the Poisson models may have grave difficulty in experiments with larger display sizes. The reason is basically the same that disconfirmed Rumelhart's (1970) Multicomponent Model: larger display sizes will impose asymptotic accuracy bounds less than 1.0. The Poisson models predict that accuracy in even the less favored locations will eventually go to $P(C_i) = 1$. If other experimental conditions with high workload continue to produce dependencies, it may be necessary to combine features of Townsend's (1981) Bounded Performance

⁷In principle, it is possible to titrate an exact amount of positive (or negative) dependence by varying the amount of capacity to be reallocated (see Townsend and Ashby, 1983, pp. 68-73). In the absence of a principled model, such an exercise was deemed to be of limited value.

Model with some form of positive dependency structure such as the Poisson or variable attention mechanism.

The positive dependencies observed in the data are not consistent with fixed-size buffer models, such as that proposed by Shibuya & Bundesen (1988) or Townsend (1981). The fixed-size buffer produces negative dependencies, because if one item is obtained, it leaves less room in the memory buffer for future items. It would be interesting to examine explicit dependency tests performed on their data, whose model fits may have implicitly tested dependencies.

Finally, and perhaps most significantly, the present analyses demonstrate that inter-item dependencies in the present data disconfirmed the assumption of independence between items processed from a multi-item display and thereby brought into question a fundamental assumption of what continues to be a modal information processing conception of whole report behavior. More experimental research will be required to tease out the conditions under which dependencies do or do not manifest themselves and what implications lie ahead for finer grained models of whole report processing. The present data and tests suggest that inter-item dependencies can provide strong model discrimination and should lead to a better articulated model of whole report processing.

References

- Atkinson, R.C. (1963). A variable sensitivity theory of signal detection. *Psychological Review*, 70, 91-106.
- Atkinson, R. C., Holmgren, J.E., & Juola, J.F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception and Psychophysics*, 6, 321-326.
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97, 523-547.
- Bundesen, C. (1993). The relationship between independent race models and Luce's choice axiom. *Journal of Mathematical Psychology*, 37, 446-471.
- Bundesen, C. (1987). Visual attention: Race models for selection from multielement displays. *Psychological Research*, 49, 113-121.
- Bundesen, C. & Pedersen, L. F. (1984). Measuring efficiency of selection from briefly exposed visual displays: A model for partial report. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 329-339.
- Busey, T.A. and Loftus, G.R. (1994). Sensory and cognitive components of visual information processing. *Psychological Review*, 101, 446-469.
- Gegenfurtner, K. & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 845-866.
- Hughes, H.C. & Townsend, J.T. (1998). Varieties of binocular interaction in human vision. *Psychological Science*, 9, 53-60.
- Kahneman, D. K. (1973). *Attention and effort*. Englewood Cliffs: Prentice Hall.
- Kullback, S. (1968). *Information theory and statistics*. New York: Dover.
- Loftus, G. R., and Busey, T.A. (1992). Multidimensional models and iconic decay: Reply to Di Lollo and Dixon. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 556-561.
- Loftus, G. R., Busey, T.A. & Senders, J. (1993). Providing a sensory basis for models of visual information acquisition. *Perception & Psychophysics*, 54, 535-554.

- Loftus, G. R., Duncan, J. & Gerhig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 530-549.
- Loftus, G. R., & Ruthruff, E. (1994). A theory of visual information acquisition and visual memory with special application to intensity-duration tradeoffs. *Journal of Experimental Psychology: Human Perception & Performance*, 20, 33-49.
- Luce, R.D. (1959). On the possible psychophysical laws. *Psychological Review*, 66, 81-95.
- Massaro, D.W. (1970). Perceptual processes and forgetting in memory tasks. *Psychological Review*, 77, 557-567.
- McGill, W.J. (1963). Stochastic latency mechanisms. In Luce, R. D., Bush, R.R. & Galanter, E. (Eds.) *Handbook of Mathematical Psychology*. New York: John Wiley & Sons, Inc.
- Norman, D.A. (1964). Sensory thresholds, response biases, and the neural quantum theory. *Journal of Mathematical Psychology*, 1, 88-120.
- Rumelhart, D.E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 7, 191-218.
- Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 591-600.
- Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs*, 74, 1-29.
- Sperling, G. (1967). Successive approximations to a model for short-term memory. *Acta Psychologica*, 27, 285-292.
- Townsend, J.T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, 10, 161-163.
- 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*. Hillsdale: NJ: Erlbaum Associates. Inc. pp. 133-168.

- Townsend, J. T. (1981). Some characteristics of visual whole report behavior. *Acta Psychologica*, 47, 149-173.
- Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan and F. Restle (Eds.), *Cognitive Theory Vol III* (pp. 200-239). Hillsdale, NJ: Erlbaum Associates.
- Townsend, J. T. & Asby F. G. (1983). *Stochastic Modeling of Elementary Psychological Processes*. Cambridge: Cambridge University Press.
- Townsend, J. T., Hu, G. G. & Evans, R. J. (1984). Modeling feature perception in brief displays with evidence for positive interdependencies. *Perception & Psychophysics*, 36, 35-49.
- Townsend, J. T., Hu, G.G. & Kadlec (1988). Feature sensitivity, bias, and interdependencies as a function of energy and payoffs. *Perception & Psychophysics*, 43, 575-591.
- Van Zandt, T. & Ratcliff, R. (1995). Statistical mimicking of reaction time data: Single-process models, parameter variability and mixtures. *Psychonomics Bulletin and Review*, 2, 20-54.

Appendix A- Capacity and Inter-Item Dependencies

The integral role of capacity in RT models can be subtle. Capacity is typically measured across variations in load (e.g., Kahneman, 1973; Townsend & Ashby, 1978). However, capacity, realized as stochastic structure in a model, can also lead to other kinds of effects within a fixed load condition. For instance, as observed earlier, the Fixed Sample Size model predicts negative dependency due to the fixed number of slots in a buffer as apparently does the Shibuya and Bundesen (1993) model which follows an independent processing stage with a fixed-size memory buffer. The Fixed and Weighted Path Poisson Models are presently mute with regard to capacity in a direct sense, because they have not been constructed to make predictions when load is varied, in either the serial or parallel interpretations. However, they predict positive dependencies as noted earlier and as will be interpreted below. With regard to load variations, if we formed a Poisson standard serial model (Townsend & Ashby, 1983, pp. 80-81), then capacity is unlimited at the individual element level but obviously limited at, say, the exhaustive processing level because RT increases dramatically as n increases. On the other hand, suppose that a parallel processing distribution was assumed to be Poisson with the overall process rate, say V held constant as load changed. Then, capacity would be limited at the individual element level, since the overall rate, say V , would be split up among all the n elements, and then reallocated among the remaining ones, as processing went along. This model has long been known to be mathematically equivalent to the standard serial Poisson Model (Atkinson, Holmgren & Juola, 1969; Townsend, 1971).

Serial models can, with the proper stochastic structure, predict positive dependencies, which may seem somewhat counter-intuitive. For example, one might assume that because only a single item is processed at once, the system has a limited capacity and would therefore predict negative dependencies. However, this logic is incorrect. The serial models we applied in this study are themselves Poisson standard serial models (Townsend, 1974) which possess exponential processing distributions on item processing times, where the rates are constant over items, location, processing position and load n . These models predict positive inter-item dependencies (e.g., Townsend & Ashby, 1983, Chapter 4). The easiest way to comprehend the positive dependency is to analyze the situation within the parallel interpretation of the standard serial Poisson model, which allows reallocation of available resources to unfinished items once an item has finished processing (also see Townsend & Ashby, 1983, pp. 80-81, for more detail). Since the models are equivalent, they must predict the same dependencies. The probability of obtaining

item "b" given "a" is obtained, provides a measure of the inter-item dependencies. Compare the probability of obtaining item "b" knowing that item "a" was obtained, relative to the probability of obtaining "b" without knowing whether "a" is finished (which is the marginal probability for "b"). In the first case, since "a" is done, "b" must have spent some time at its higher reallocated rate and thus has higher (than its marginal) probability of being finished as well. The fact that "b" received additional resources once "a" terminates gives the two items a positive inter-item dependency. A strict serial interpretation follows directly from this example, since in this case the processing resources are directed to the individual items in an all-or-none fashion, and item "b" receives no processing until item "a" has terminated. To see how a limited capacity independent model can be limited (in fact fixed) capacity, consider Parallel Model #3 (p.85) in Townsend and Ashby (1983). That and other models on capacity (pp. 76-91) help make explicit the relations between capacity and dependency.

Appendix B- Serial and Parallel Model Mimicking

The Weighted Path Serial Poisson Model is mathematically equivalent to a parallel model with the following characteristics. Assume that when processing starts, processing begins at each position with an exponential distribution and at rates $V_{11}, V_{21}, V_{31}, V_{41}$, where the first subscript denotes physical location and the second, stage 1. These are constrained such that

$$\sum_{i=1}^4 V_{i1} = \lambda \tag{Eq A1}$$

where λ represents the serial processing parameter. Set

$$P(\text{location } i \text{ is processed first}) = \frac{V_{i1}}{\sum_{j=1}^4 V_{j1}} = P_i \tag{Eq A2}$$

in the serial Weighted Path Model. Suppose the item in stimulus location j was processed first. Then we assume that all processing (capacity) allotted to j is reallocated, after j 's completion, to the remaining items in such a way that the relative magnitudes are the same as in stage 1. This implies, that, say,

$$P(i \text{ is } 2^{nd} | j \text{ is } 1^{st}) = \frac{V_{i2}}{\sum_{\substack{k=1 \\ k \neq j}}^4 V_{k2}} = \frac{P_i}{\sum_{\substack{k=1 \\ k \neq j}}^4 P_k} \tag{Eq A3}$$

and that $\sum_{\substack{k=1 \\ k \neq j}}^4 V_{k2} = \lambda$. This model is a special case of the Reallocation Parallel Model which is equivalent to

the Poisson standard serial model exponential random path serial model (Townsend, 1974; Townsend & Ashby, 1983, pp. 88-89). The serial model assumes that processing on each item is unaffected by the number of items in the display, and thus the serial interpretation is designated as Unlimited Capacity in Figure 1, as described in the definition of the Fixed Path Poisson model in the text. The parallel model is made equivalent to the serial model by making provisions for a proper reduction in overall processing capacity as the number of items (= load) increases (e.g. Townsend, 1974). In the present study, load was held constant. Note that in the parallel interpretation, the processing rate allocated to each item in the display equals λ/n , which is affected by the number of items in the display, and thus this model is

designated as Limited Capacity in Figure 1. The parallel version of this model is clearly equivalent to the Weighted Path Serial Model.

We will skirt detailed descriptions of other mimicking models but namely note that the Variable Attention Model depicted as parallel could be presented as serial. All that is necessary is to provide the serial model that mimics a parallel independent model to also possess two distinctive sets of processing rates (e.g. see Townsend & Ashby, 1983, pp. 82-83) corresponding to the switches in attention within the Variable Attention Model.

On the other hand, the Fixed Path Poisson Model cannot be perfectly mimicked by any parallel model because any parallel model attempting to mimic this model has all its rate parameters but one going to zero at any particular stage. Obviously, a parallel model could nevertheless closely approximate that model.

Although certain of these models have more compelling interpretations in one of the architectures, it is their dependency structure that is being most strikingly tested in this study.

Author Note

While conducting this research Thomas Busey was supported by a two-year National Institute of Mental Health fellowship. Dr. Townsend was supported in part by a National Science Foundation Grant #9112813. The authors gratefully acknowledge Dasha Kinelovsky for her assistance with the joint probability predictions.

Glossary

PHYSICAL ATTRIBUTES		UNITS	
d	Stimulus duration	msec	
t	Time since stimulus onset	msec	
TERMS COMMON TO ALL MODELS			
$P(C_i)$	Probability of acquiring the digit in stimulus location i .	probability	
$P(C_j \bar{C}_i)$	Probability of correctly acquiring an item to the right of the first error, where i was the location of the first error	probability	
X_i	Random variable that determines whether digit in location i ($i=1..4$) is correct ($X_i = 1$) or incorrect ($X_i = 0$)	0 or 1	
g	Guessing rate (equal to 0.1 in the current data)	proportion	
BOUNDED PERFORMANCE MODEL		UNITS	
$I(\infty)$	Asymptotic level of information	proportion	
t_0	Pre-processing delay	msec	
V	Rate of growth	1/msec	
ξ	Amount of available capacity	number of items (not necessarily an integer)	
a_i	Proportion of capacity allocated to item i	proportion	
INDEPENDENT SAMPLING MODEL			
c	Estimated slope inverse	msec	
L	Estimated duration-axis intercept	msec	
Y	Estimated performance asymptote	proportion	
FIXED AND WEIGHTED PATH POISSON MODELS		UNITS	DEPENDS ON
$P(j, t, t_0)$	Poisson probability that j letters have been acquired by time t	probability	j, t, t_0
λ	Processing rate of the Poisson process	items/msec	
$u_{i,n}$	Probability that stimulus location i is processed in position n of the Poisson process	probability	$w_{i..1}$
$w_{i..1}$	Weight assigned to locations $i..1$ when computing the probability of taking a particular path through 4 stimulus locations	probability	
[J]	One of 24 possible paths through the four stimulus locations, $i=1..4$	array of locations	

$\phi([J_i])$	Probability of taking path $[J_i]$	probability	$w_{i..1}$
Y_i	Random variable that determines whether digit in processing position i ($i=1..4$) is correct ($X_i = 1$) or incorrect ($X_i = 0$)	0 or 1	
VARIABLE ATTENTION MODEL		UNITS	DEPENDS ON
μ	mean of the binomial function determining the distribution of attentional states; determines the skewness of the binomial	proportion	
σ	standard deviation of the spread of attention within the binomial distribution of states. Determines how much the low and high attention states deviate from the average rate λ .	proportion of λ	
a	One of five attentional states.	integer	
$\alpha(\mathbf{a})$	Probability of being in attentional state a , $\mathbf{a}=\{0..4\}$	probability	μ, \mathbf{a}
pace _{i}	Processing pace associated with stimulus location i	msec	$\mu, \alpha(\mathbf{a}), \sigma$

Tables

Observer	1/c	L	λ	t_0	w ₁	w ₂	w ₃	w ₄
TB	117.6	84.6	0.0275	91.3	0.587	0.308	0.0758	0.0290
SS	120.1	72.5	0.0261	77.4	0.273	0.479	0.228	0.0203
EF	32.4	62.6	0.0717	61.2	0.281	0.535	0.156	0.0285

Table 1. Model parameters for the Independent Sampling Model (1/c and L), the Fixed Path Poisson Model (λ and t_0) and the Weighted Path Poisson Model (w₁...w₄). The Weighted Path Poisson Model, described in a subsequent section, uses the λ and t_0 parameters estimated from the Fixed Path Poisson Model applied to the marginal data of Figure 2. See Figure 2 for RMSE's associated with each model fit.

Stimulus Duration (ms)	Observer TB	Observer SS	Observer EF
40	3.18	7.94	8.70
50	2.79	9.90	9.11
63	3.70	5.13	11.8
80	12.4	6.24	22.4
100	6.00	22.9	23.2
126	30.6	33.6	13.2
159	11.4	31.6	0.62
200	21.8	21.0	0.00

Table 2. Independence statistics computed from Eq 4 for three observers at 8 stimulus durations. Values that exceed 19.67 demonstrate evidence of inter-item dependencies. Stimulus durations of less than 80 ms for TB and SS and 63 ms for EF produce only chance performance, which will not contain dependencies. Longer stimulus durations produce evidence of inter-item dependencies, although Observer EF's performance reaches 1.0 at long stimulus durations and thus she cannot show inter-item dependencies at these durations.

Model	Core Assumptions	Predicts Positive Inter-item Dependencies	Rank Order Accounts of Log-Likelihood
Independent Sampling Model	Processing proceeds in parallel and independently on all four stimulus locations. The processing time is exponentially distributed with a pre-processing delay. Each stimulus location has a separate processing rate.	no	3
Fixed-Path Poisson Model	Processing occurs serially in a left-to-right order. The processing times for each location are exponentially distributed; the overall process has a pre-processing delay.	yes	-
Weighted Path Poisson Model	Identical to the Fixed-Path Poisson Model, with the exception that processing can occur through any one of the 24 possible processing paths through the 4 stimulus locations.	yes	2
Variable Attention Model	As with the Independent Sampling Model, processing occurs in parallel with separate rates for the four stimulus locations. Attention is assumed to vary across trials (but remain fixed within a trial). This attention is assumed to be pseudo-normally distributed. The value of the attention parameter modulates the processing rate for each location by a multiplicative amount. Thus in a relatively high attentional state, all four processing rates will be high.	yes	1

Table 3. Descriptions of the assumptions underlying the 4 models tested in the current work. The core assumptions listed describe the typical interpretations of the model; see Appendix B for alternative interpretations which provide isomorphic models.

Observer TB

Stimulus Duration (ms)	Ave[P(C _i)]	Ave[P(C _i C _j)-P(C _i)]	Fixed Path Poisson:	Weighted Path Poisson:	Variable Attention:
80	0.130	0.080	0.026	0.001	0.002
100	0.139	0.019	0.043	0.008	0.024
126	0.301	0.171	0.119	0.074	0.082
159	0.588	-0.003	0.140	0.107	0.051
200	0.676	0.065	0.096	0.080	0.031
RMSE			0.027	0.040	0.035

Observer SS

Stimulus Duration (ms)	Ave[P(C _i)]	Ave[P(C _i C _j)-P(C _i)]	Fixed Path Poisson:	Weighted Path Poisson:	Variable Attention:
80	0.102	-0.038	0.006	-0.002	0.002
100	0.245	0.068	0.071	0.031	0.021
126	0.431	0.072	0.134	0.078	0.068
159	0.574	0.063	0.132	0.096	0.075
200	0.713	0.058	0.082	0.068	0.061
RMSE			0.022	0.018	0.018

Observer EF

Stimulus Duration (ms)	Ave[P(C _i)]	Ave[P(C _i C _j)-P(C _i)]	Fixed Path Poisson:	Weighted Path Poisson:	Variable Attention:
63	0.132	0.125	0.004	0.007	-0.005
80	0.417	0.119	0.140	0.083	0.098
100	0.757	0.051	0.088	0.085	0.026
126	0.889	0.014	0.023	0.031	0.007
159	0.951	-0.001	0.003	0.006	0.001
200	0.972	0.000	0.000	0.000	0.000
RMSE			0.049	0.048	0.053

Table 4. Inter-Item Dependency Statistic Ave[P(C_i|C_j)-P(C_i)], compared with the predictions of the Fixed-Path Poisson Model, and the Weighted Path Poisson Model, which is described in a subsequent section. The Independent Sampling Model predicts that the values for Ave[P(C_i|C_j)-P(C_i)] will be 0.0. A sign test aggregating all stimulus durations for the three participants demonstrates that out of 16 computed dependencies, 12 are positive, which is significant at $\alpha = 0.05$. This disconfirms the Independent Sampling Model. The Fixed Path Poisson Model can account for these dependencies, while the Weighted Path Poisson and Variable Attention Models can as well.

Dur.	Observer SS				Observer EF				Observer TB			
40	28	5	6	0	17	2	5	0	36	2	1	0
	5	1	1	1	5	0	4	0	5	1	0	0
	4	1	1	0	1	1	1	0	7	0	0	0
	0	0	1	0	0	0	0	0	2	0	0	0
50	42	1	2	0	28	2	2	0	41	4	4	0
	5	1	0	0	2	0	0	0	1	0	0	0
	2	0	0	0	1	0	0	0	3	0	1	0
	0	1	0	0	0	1	0	0	0	0	0	0
63	37	2	4	0	23	1	2	0	38	1	2	0
	5	1	0	0	3	1	2	0	3	0	0	0
	5	0	0	0	2	0	1	1	9	1	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
80	34	5	4	0	5	3	0	0	31	9	3	1
	3	1	1	0	5	10	0	0	5	0	0	0
	6	0	0	0	3	3	1	0	3	1	0	0
	0	0	0	0	0	3	0	3	0	0	0	1
100	19	11	6	0	0	0	0	1	31	3	2	0
	1	3	1	0	3	3	0	1	8	2	1	0
	4	5	0	0	1	3	0	1	3	0	1	0
	0	3	0	1	0	7	0	16	3	0	0	0
126	12	1	0	0	0	0	1	0	15	10	0	1
	3	9	2	1	0	2	0	2	11	8	1	0
	4	3	3	3	0	0	0	1	0	0	2	0
	1	10	1	1	0	6	0	24	1	2	0	3
150	5	2	1	0	0	0	0	0	2	1	2	0
	3	6	3	0	0	0	0	0	3	10	3	6
	1	6	0	0	0	0	0	2	0	4	0	1
	1	18	1	7	0	5	0	29	5	9	3	5
200	3	1	0	1	0	0	0	0	1	2	0	0
	1	4	0	3	0	0	0	0	4	13	0	4
	0	4	0	2	0	0	0	0	1	0	0	1
	3	15	0	17	0	4	0	32	0	13	1	14

Table 5. Counts of joint events for 3 Observers for 8 stimulus durations. The left-most entry is the left-most stimulus position in the display.

Key:

- www wwwc wwww
- cwww ccww cwww
- wwwc wwww wwww
- cwww ccww cwww

Observer	Number of Durations Analyzed	Degrees of Freedom (5 free params)	Independence Log-Likelihood (5 free parameters)	Weighted-Path Poisson Log-Likelihood (5 free parameters)	Critical X^2 value $\alpha = 0.01$ (5 free params)	Variable Attention Log-Likelihood (7 free params)	Critical X^2 value $\alpha = 0.01$ (7 free params)
TB	4	55	85.9 *	85.17 *	82.3	71.30	79.8
SS	5	70	125.7 *	106.32 *	100.4	91.04	98.0
EF	4	55	90.4 *	79.76	82.3	75.35	79.8

Table 6. Log-likelihood values for the three models for the above-chance and below-ceiling stimulus durations. The log-likelihood values have been multiplied by 2.0 to allow comparison with the critical X^2 value. Starred items indicate that the model was rejected for that observer. Of the three models, the Variable Attention Model performs the best, even when the values are corrected for the number of free parameters via the AIC conversions, which adds twice the number of free parameters to each log-likelihood value (not shown). However, the model performs best by using extremely large values for the variance of the variable attention across trials.

Observer	L	1/c ₁	1/c ₂	1/c ₃	1/c ₄	σ	p
TB- 5 state	98.4	43.76	63.13	177.81	370.92	.479	.496
SS- 5 state	88.04	309.5	217.89	359.84	1758.7	.442	.123
EF- 5 state	72.6	19.57	12.57	25.7	58.61	.490*	.431

Table 7. Model parameters for the Variable Attention Model. In the case of observer EF, σ was fixed at .490 in order to avoid negative processing paces.

Figure Headings

Figure 1. Taxonomy of whole-report models, including processing assumptions that affect capacity and inter-item dependencies.

Figure 2. Performance curves for three observers from Loftus, Busey & Senders (1993). Curves represent the best fits for the Independent Sampling and Fixed-Path Poisson models. Although the two models are structurally quite different, they make very similar predictions for the marginal data, suggesting that other analyses such as the inter-item dependencies are required to discriminate between models.

Figure 3. **Left Panels.** Plots of the statistic $P(C_i | \text{previous error})$ for 8 stimulus durations for Observers TB and SS in which more than 10 trials contribute to each data point. Observer EF generated too few errors to make this analysis reliable. The Independent Sampling Model predicts no differences between the curves, only that they should increase as a function of exposure duration at a rate equal to the marginal performance curves shown in Figure 2. The Fixed Path Poisson Model assumes that the curves should remain constant at the guessing rate of 0.1. **Right Panels.** Fit of the Weighted Path Poisson Model to the $P(C_i | \text{previous error})$ data.

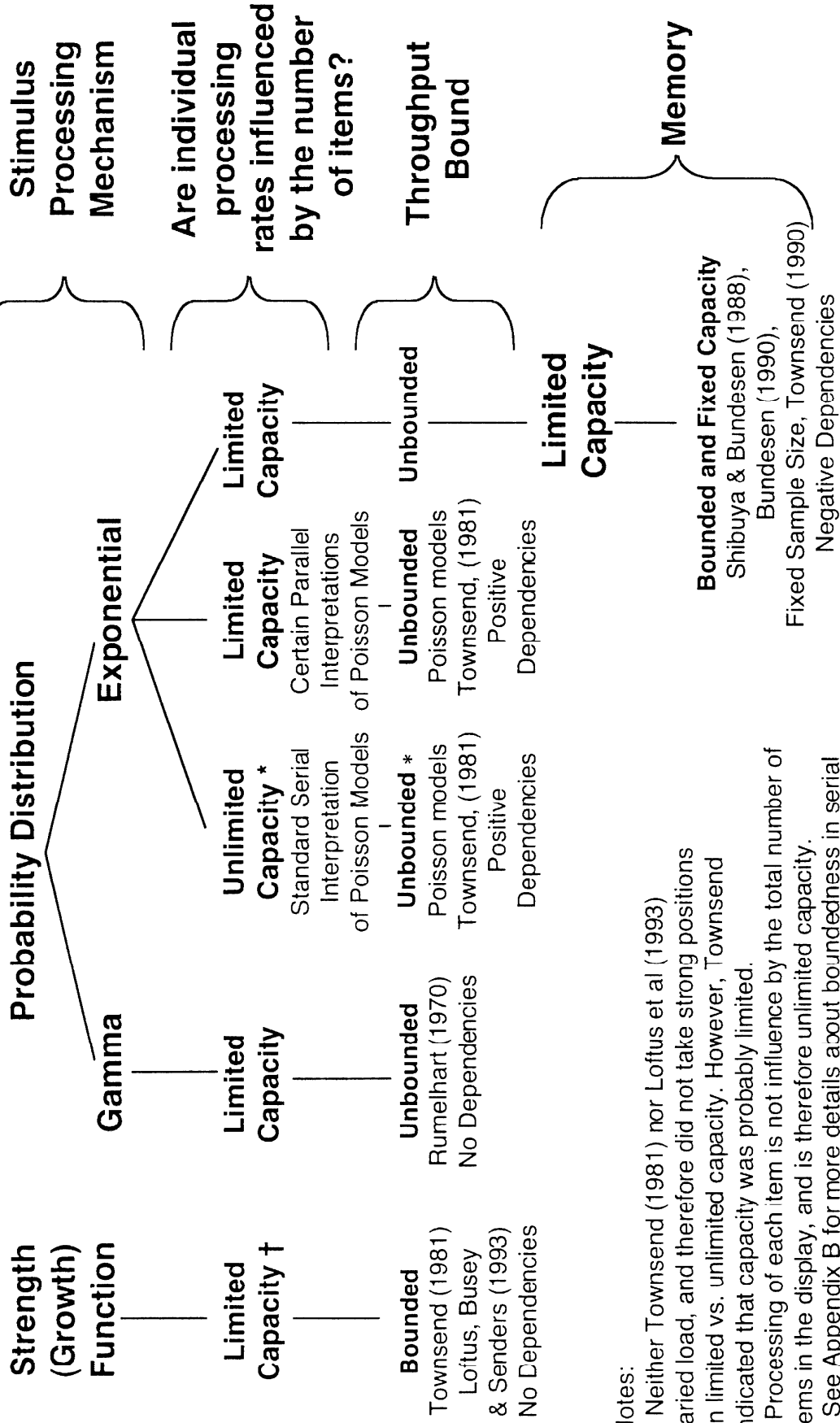
Figure 4. The inter-item dependency statistics $\text{Ave}[P(C_i | C_j) - P(C_i)]$ plotted against $\text{Ave}[P(C_i)]$. Time is an implicit parameter used to generate the predicted curves. Lower curve is the Fixed Path Poisson Model prediction, and the Independent Sampling Model prediction falls along the abscissa. These fits are generated from the parameters fit to the marginal data, and thus these fits have no free parameters. Data for three subjects from Loftus et al (1993) are shown as separate points. The Fixed Path Poisson Model prediction is generated from parameters fit to averaged marginal data, and is shown to demonstrate the general effects of the model. Predictions and computed statistics for individual observers are found in Table 3.

Figure 5. Probability of getting digit C correct given exactly k other digits were correct, averaged over position $i = 1$ to 4, plotted for $k = 0$ to 3. Top row: predictions for the Independent Sampling Model. Bottom row: predictions for the Fixed Path Poisson Model. Both sets of model predictions were generated by fitting each model to the observed marginal probability data using two free parameters for each model, and then applying these parameters to generate the dependency predictions. Thus each set of eight curves for each model is based on two estimated parameters, although the parameters are estimated from the marginal data.

Figure 6. Probability of getting digit C correct given exactly k other digits were correct, averaged over position $i = 1$ to 4, plotted for $k = 0$ to 3, along with the predictions for the Weighted Path Poisson and Variable Attention Models.

Figure 7. Autocorrelation functions along with 95% confidence bands for the correlation at each lag.

Model Mechanism



Stimulus Processing Mechanism

Are individual processing rates influenced by the number of items?

Throughput Bound

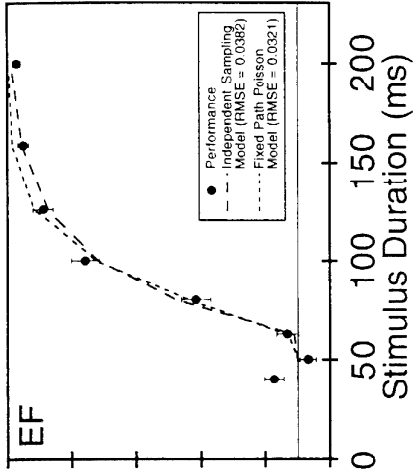
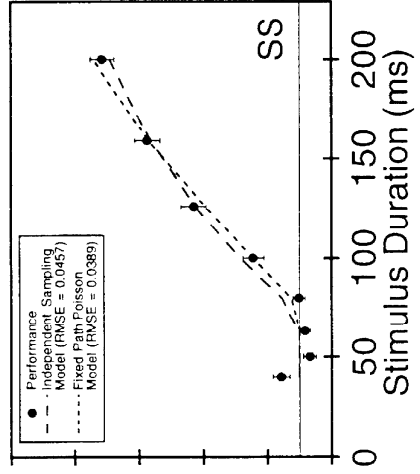
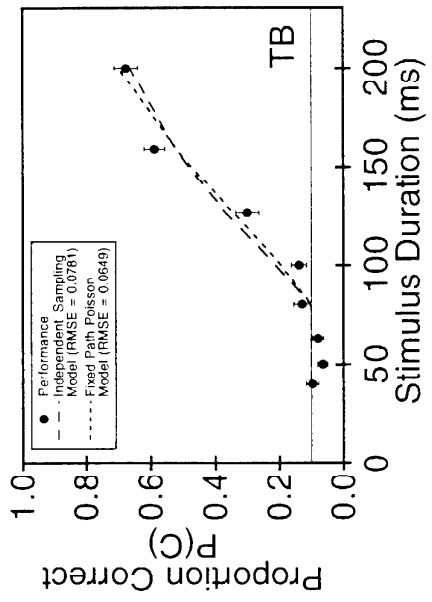
Memory

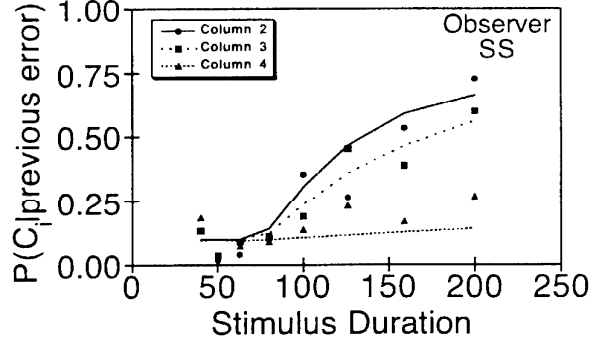
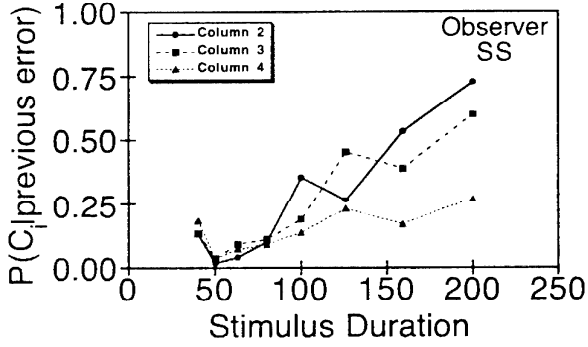
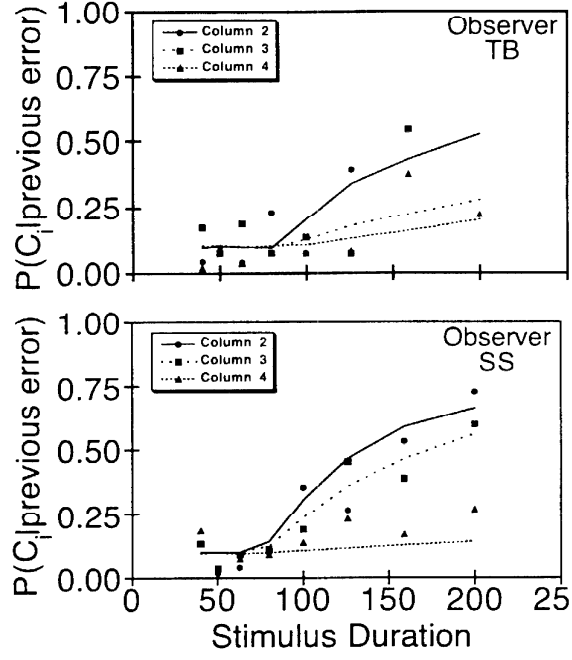
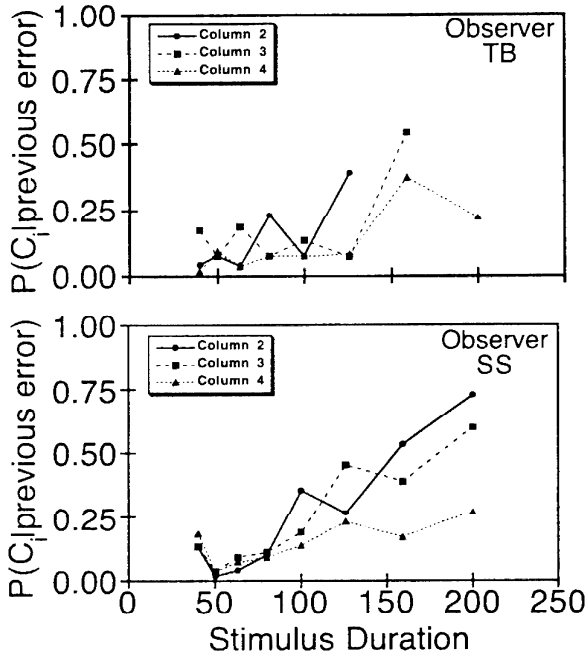
Limited Capacity

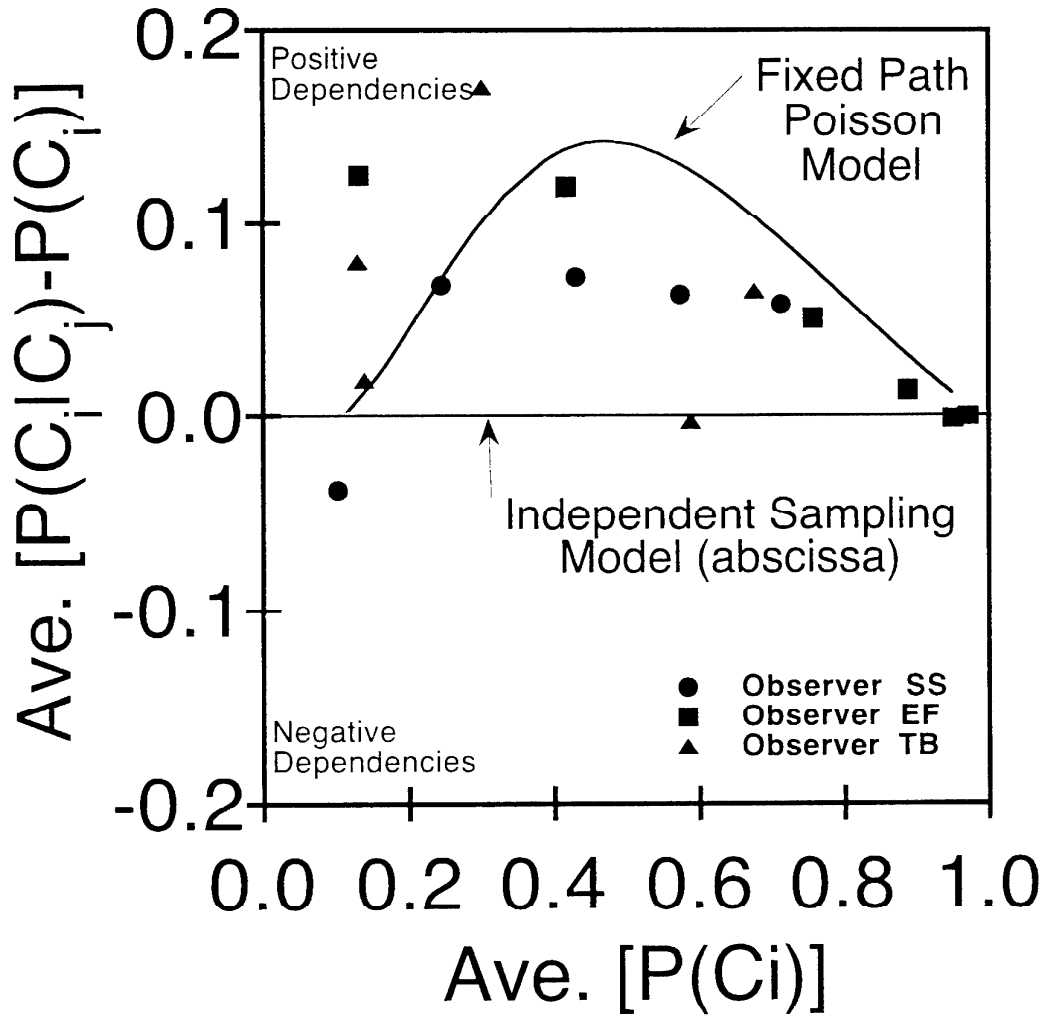
- Bounded and Fixed Capacity**
 - Shibuya & Bundesen (1988), Bundesen (1990),
- Fixed Sample Size, Townsend (1990)**
- Negative Dependencies**

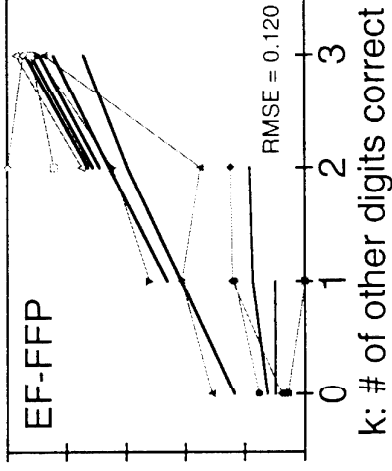
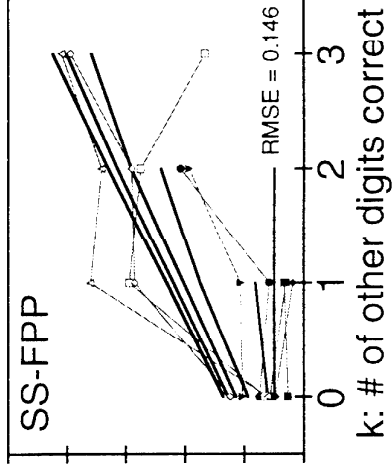
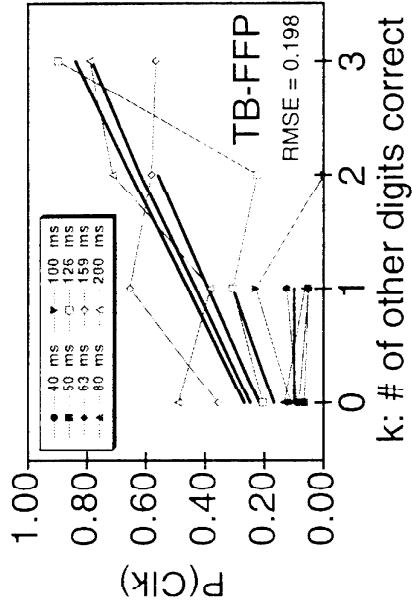
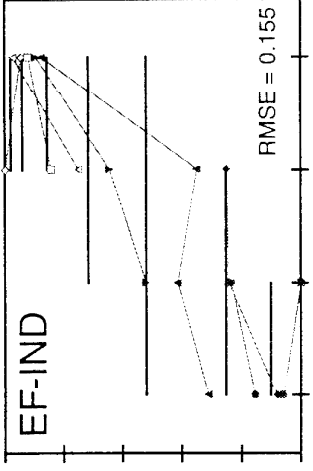
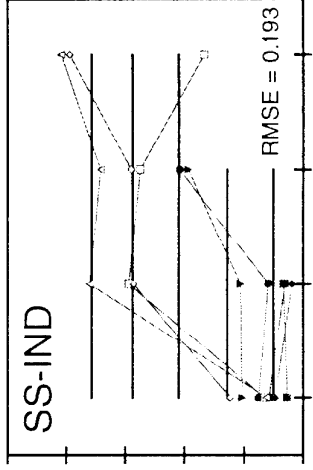
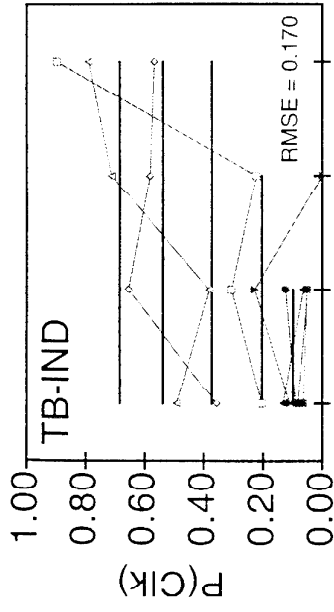
Notes:

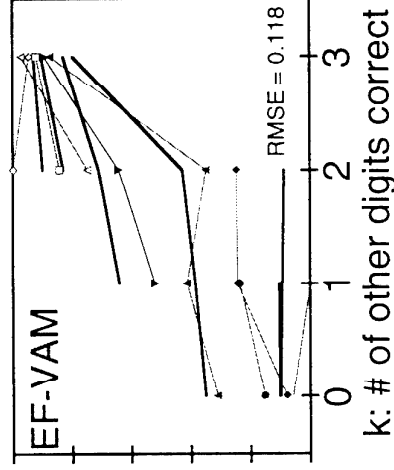
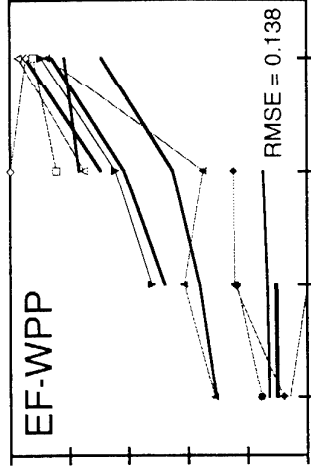
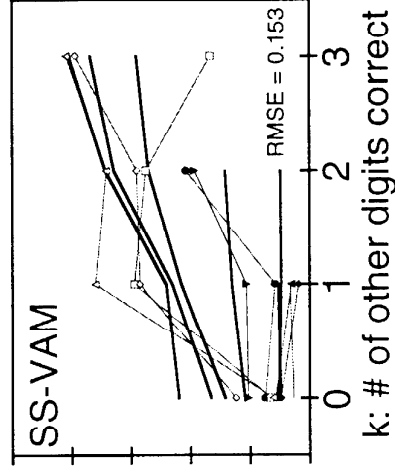
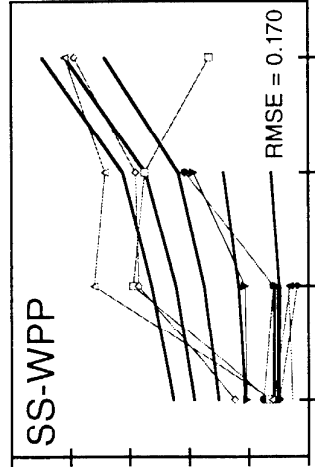
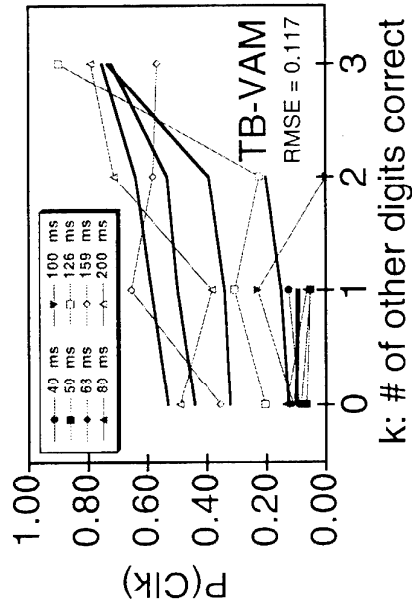
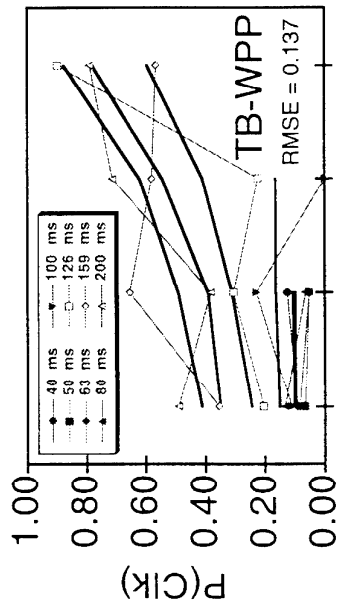
- † Neither Townsend (1981) nor Loftus et al (1993) varied load, and therefore did not take strong positions on limited vs. unlimited capacity. However, Townsend indicated that capacity was probably limited.
- * Processing of each item is not influenced by the total number of items in the display, and is therefore unlimited capacity.
- * See Appendix B for more details about boundedness in serial models.

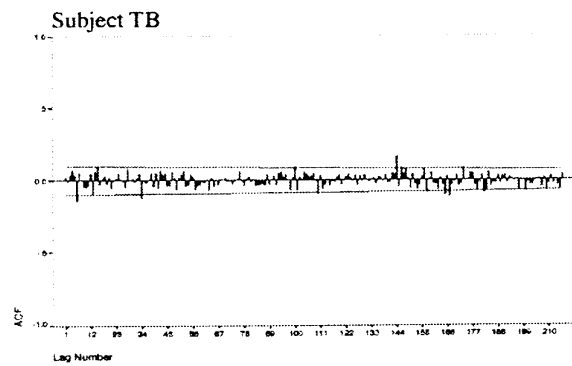
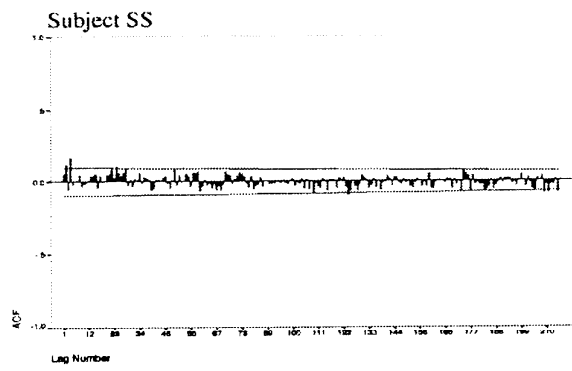
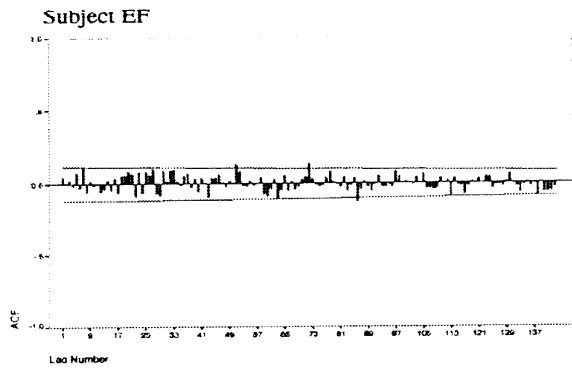












7

