Behav Res (2012) 44:1148–1156
DOI 10.3758/s13428-012-0202-4

Numerical predictions for serial, parallel, and coactive logical rule-based models of categorization response time

Daniel R. Little

Published online: 21 April 2012

Abstract Recent theoretical advances in theories of categorization response times have made it possible to differentiate mental architectures that specify how processes occurring over several information-processing channels are combined (e.g., in serial or in parallel). This article introduces the numerical computations necessary to generate predictions for a class of logical rule-based models that have recently been used to account for speeded perceptual categorization judgments (Fifić, M., Little, D.R. & Nosofsky, R. M. . Psychological Review, 117:309–348, 2010).

Keywords Response times · Information processing architecture · Serial/parallel processing

Recent work in perceptual categorization has built upon the classic idea that people learn and represent categories by using simple, logical rules (Bourne, 1970; Bruner, Goodnow, & Austin, 1956; Levine, 1975; Trabasso & Bower, 1968). That is, a category might be represented by a conjunction (AND) or a disjunction (OR) of rules applied to different stimulus dimensions. Fifić, Little, and Nosofsky, (2010) and Little, Nosofsky, and Denton (2011) proposed a set of logical-rule models capable of explaining and generating predictions for the full time course of categorization decision making. By combining sequential-sampling models (Busemeyer, 1985; Ratcliff, 1978) with mental-architecture models (Schweickert, 1992; Sternberg, 1969; Townsend, 1984; Townsend & Nozawa, 1995; Townsend & Wenger, 2004a, 2004b) of response time (RT) in an integrated framework, the logical-rule models make predictions at the level of full RT distributions. This approach has achieved considerable success in answering fundamental questions about about whether multiple stimulus dimensions are processed sequentially in a serial fashion or simultaneously in parallel, or are pooled into a single, coactive processing channel (Houpt & Townsend, 2011; Miller, 1982; Townsend & Wenger, 2004a, 2004b; see also Eidels, Houpt, Altieri, Pei, & Townsend, 2011; Palmer & McLean, 1995; Thornton & Gilden, 2007).

The logical-rule models assume that when presented with a multidimensional perceptual stimulus, subjects make independent decisions about the category outcomes predicted by the values of each of the dimensions. Each independent decision is instantiated as a separate random-walk process (Busemeyer, 1985; Luce, 1986). The independent decisions are then combined according to rules specified by alternative mental architectures that define the category structure. For example, consider the category space shown in Fig. 1. In this category space, the stimuli vary along two continuous dimensions, x and y, with three values per dimension combined orthogonally to yield nine distinct stimuli. The stimuli above and to the right of the dotted category boundary are members of Category A; the remaining stimuli are members of Category B.

Category A is defined by a conjunction of two rules. That is, a stimulus is a member of Category A if it has a value greater than or equal to x_1 on dimension x AND greater than or equal to y_1 on dimension y. Consequently, an exhaustive stopping rule is necessary for combining independent decisions made about the values of x and y in order to classify a stimulus as a member of Category A. By contrast, a stimulus is a member of Category B if it satisfies a disjunctive rule; that is, the stimulus has a value less than x_1 on dimension x OR less

Electronic supplementary material The online version of this article (doi:10.3758/s13428-012-0202-4) contains supplementary material, which is available to authorized users.

D. R. Little (✉)
Psychological Sciences, University of Melbourne,
Parkville, Victoria 3053, Australia
e-mail: daniel.little@unimelb.edu.au

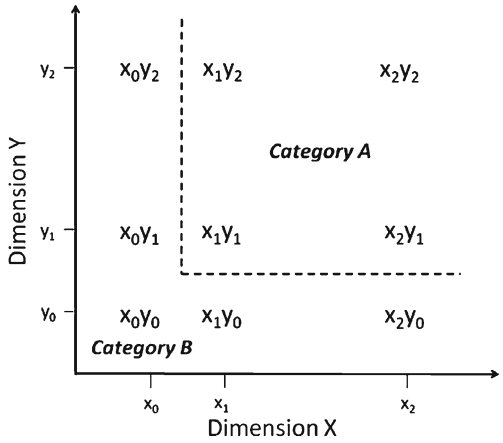

Fig. 1 Schematic illustration of the category space used for testing the logical-rule models. The stimuli are composed of two dimensions, *x* and *y*, with three values per dimension, combined orthogonally to produce the nine members of the stimulus set. The stimuli in the upper-right quadrant of the space are the members of Category A, whereas the remaining stimuli are the members of Category B

than y_1 on dimension *y*. Hence, a self-terminating stopping rule is appropriate for combining the features of the Category B stimuli. An exhaustive stopping rule might still be applied to make Category B decisions; however, unlike for Category A, an exhaustive stopping rule is not mandated by the category space. These stopping rules may be implemented either serially, by processing the stimulus dimensions one at a time, or in parallel, by processing the stimulus dimensions simultaneously. Alternatively, multiple stimulus dimensions might be pooled into a single, coactive processing channel (Houpt & Townsend, 2011; Miller, 1982; Townsend & Nozawa, 1995).

When generating predictions for serial and parallel combinations of sequential sampling models, simulation methods have been utilized to generate predictions in order to fit the models to the observed RTs (Fifić et al., 2010; Little et al., 2011; Little, Nosofsky, Donkin, & Denton, 2012; Thornton & Gilden, 2007). One potential difficulty with developing fast, numerical predictions for the RT distributions is that the serial and parallel models both involve a combination of more than one random walk. A further complication is that to fit distributions of the overall RTs, the distribution of decision times must be convolved with distributions of residual nondecision times. Though software exists to fit and compute the predictions of single-channel sequential-sampling models (e.g., fast-DM, Voss & Voss, 2008; DMAT, This citation should be Vandekerckhove & Tuerlinckx, 2008; or EZ-diffusion, Wagenmakers, van der Maas, & Grasman, 2007), combining multiple predictions of these single-channel models is not straightforward. Furthermore, these tools are designed to fit a continuous-time sequential-sampling model (the diffusion model; Ratcliff, 1978) rather than the discrete-time random-walk models employed in the logical-rule models (Fifić et al., 2010; see also Logan, 2002; Nosofsky & Palmeri, 1997; Thornton & Gilden, 2007). The purpose of this article is to introduce the numerical computations necessary to generate predictions for the logical rule-based models (Fifić et al., 2010; Little et al., 2011; Little et al., 2012).[1]

The present method adopts a parametric model-fitting approach, in contrast to the nonparametric results reported for serial and parallel models (Townsend & Ashby, 1983; Townsend & Nozawa, 1995; Townsend & Wenger, 2004b). There are several benefits to adopting a parametric, sequential-sampling model approach: First, the actual observed behavior may not be best captured by a purely serial or purely parallel model, but instead one might augment these processes with additional mechanisms. For example, in Little et al. (2011), an additional attention-switching mechanism was necessary to capture the time taken to switch responding from one serially processed dimension to the next. This approach allows for efficient exploration of the space of the cognitive processes and mechanisms that may be relevant to decision making. Second, this approach allows for direct comparison of a number of different models. For instance, Fifić et al. (2010) tested the predictions of eight other models, including two sequential-sampling models that are among the leading theoretical models in the field (i.e., the exemplar-based random-walk model [EBRW; Nosofsky & Palmeri, 1997] and stochastic general recognition theory [GRT; Ashby, 2000]). Third, sequential-sampling models have recently been shown to be neurally plausible models of decision making. For example, Smith (2010) recently showed that, under certain conditions, a model of neural firing converges to a sequential-sampling process that mimics the approach taken in the present report (see also Smith & Ratcliff, 2004; Smith & McKenzie, 2012). Finally, the present approach was adopted in the models of perceptual categorization presented in Fifić et al. (2010) and Little et al. (2011). The findings of those reports provide a challenge to existing models, which do not take into account the decision-making architecture. In sum, the benefits include the ease with which the models can be tested, extensibility, neural plausibility, and a relation to previously published work; however, the parametric approach taken here should be viewed as complementary to the nonparametric approach. In fact, the two approaches can provide converging evidence regarding the underlying cognitive architecture (Eidels, Donkin, Brown, & Heathcote, 2010).

In the following section, the computations for single-channel random walks are introduced, followed by the methods necessary to combine these channels into serial and parallel processes.

[1] MATLAB code for all of the computations is provided in the online supplement.

Single-channel random-walk models

In the logical-rule models, a decision about a single stimulus feature is represented by a single-channel random walk. Random-walk models are a class of sequential-sampling models in which discrete evidence accumulation drives a decision process that is completed in discrete time. Figure 2 illustrates the standard random-walk process for three example trials. Information accumulation begins at a starting point, 0, and at each time step, with probability p takes a step toward the upper boundary, $+A$, or with some probability $q = 1 - p$ takes a step toward the lower boundary, $-B$.[2] In a random walk, the boundaries may be equidistant from the starting point (i.e., $A = B$) or may be located at different distances from the starting point. In both cases, there are $m = A + B - 1$ *transient* steps between the boundaries (see, e.g., Diederich & Busemeyer, 2003). When the position of the decision process hits one of the boundaries, the process ends and the response is executed; hence, the total number of states in the random walk is $N_T = m + 1 = A + B$; that is, the total number of possible positions in the random walk is the number of transient steps, plus one for the absorbing or terminating step at either of the boundaries. In most applications, the boundaries are at different distances from the starting point; $N_A = A$ is the number of steps from the starting point to the $+A$ boundary, and $N_B = B$ is the number of steps from the starting point to the $-B$ boundary.

For a single-channel random-walk process, analytic solutions are well known for calculating the probability that the process terminates at the upper boundary or the lower boundary, along with the expected number of steps that it takes to terminate at that boundary. For example, Busemeyer and Diederich (Busemeyer & Diederich, 2010, p. 117; Diederich & Busemeyer, 2003) used a matrix representation of the random walk to derive predictions for the choice probabilities, expected decision times, and distribution of the decision times. The probability of responding with the response, R, associated with either boundary $+A$ or $-B$, starting from zero, is

$$P(R = A) = \begin{cases} \frac{N_B}{N_T}, & \text{if } p = q \\ \frac{1-\left(\frac{q}{p}\right)^{N_B}}{1-\left(\frac{q}{p}\right)^{N_T}}, & \text{if } p \neq q \end{cases} \quad (1)$$

For $P(R = B)$, replace q/p with p/q and N_B with N_A. The expected number of steps, N, in the random walk, assuming starting from zero, is given by

[2] Here we use the formulation adopted by Nosofsky and Palmeri (1997). Feller (1968, chap. 14) and Ratcliff (1978) used an alternative formulation of the random walk based on the gambler's ruin problem, which sets a lower boundary of zero, an upper boundary at A, and a starting point, Z, between 0 and A. Both formulations are equivalent.

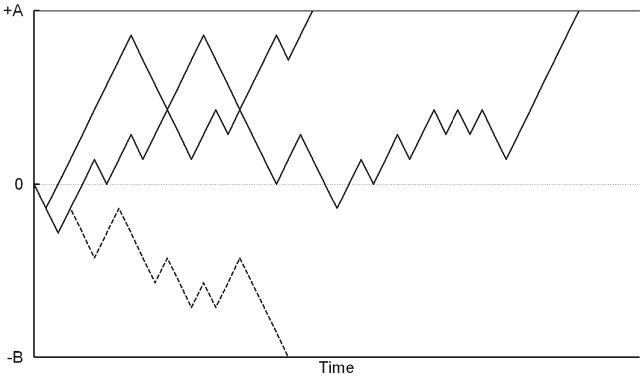

Fig. 2 Random walk example showing three possible sampling paths. At each discrete time step, the process takes a step up or down until terminating at either the upper (i.e., the solid lines) or lower boundaries (i.e., the dashed line). The boundary that the processes ends at determines the response and the total number of steps determines the time taken to make the decision

$$E(N) = \begin{cases} N_A \times N_B; & \text{if } p = q; \\ \frac{N_B}{q-p} - \frac{N_T}{q-p}\left(\frac{1-\left(\frac{q}{p}\right)^{N_B}}{1-\left(\frac{q}{p}\right)^{N_T}}\right) & \text{if } p \neq q. \end{cases} \quad (2)$$

The first-passage-time probabilities (i.e., the probabilities assigned to the possible times when the random walk first hits a boundary) for ending the random walk after N steps at boundary $+A$ (see Feller, 1968) is given in Busemeyer and Diederich (2010, chap. 4 appendix; see also the appendix of Ratcliff, 1978) as

$$f(N,\ R = A) = \frac{2^{N+1}}{N_T}\left(\sqrt{pq}\right)^{N+1}\left(\sqrt{\frac{p}{q}}\right)^{N_A} \times \sum_{j=1}^{m} \cos^N\left(\frac{\pi \cdot j}{N_T}\right) \sin\left(\frac{\pi \cdot j \cdot N_A}{N_T}\right) \sin\left(\frac{\pi \cdot j}{N_T}\right). \quad (3)$$

For $f(N, R = B)$, replace p/q with q/p and N_A with N_B. Note that the sum in Eq. 3 is over all of the possible transient positions (i.e., all of the possible distances from the $+A$ boundary). Equation 3 is not a probability distribution; summing $f(N, R = A)$ over N will give $p(+A)$. Dividing $f(N, R = A)$ by $p(+A)$ will give the conditional probability of ending at boundary A at time step N. Summing $f(N, R = A)$ and $f(N, R = B)$ at each time step will give the total probability of the random walk terminating at time step N.

Equation 3 gives the probability of ending the random walk after a number of steps, N, at the A boundary. In the logical-rule models, the number of steps is scaled to milliseconds by multiplying N by a scaling constant k.

Nondecision time

The equations above allow one to compute the decision component of the RT. In the logical-rule models, it is also assumed that an additional, *nondecision time* component is

necessary to account for processing related to stimulus encoding and response execution. This nondecision time is assumed to be log-normal with a mean μ_R and variance σ^2_R. A log-normal distribution provides an approximation to simple RT distributions because it is bounded at zero and has a positive skew (see, e.g., Luce, 1986). A log-normal nondecision time distribution, $d(x)$, can be added to the random-walk model by convolving the decision time distribution, $f(N, R = A)$, with the nondecision time distribution:

$$[f * d](N,\ R = A) = \sum_{k=0}^{N} f(k,\ R = A)d(N - k). \tag{4}$$

Determining p and q

The logical rule models incorporate the assumptions of general recognition theory (GRT; Ashby & Townsend, 1986) to determine the probability of taking a step toward the $+A$ boundary. Specifically, for a given stimulus, the models assume that the subject establishes boundaries along each dimension, as illustrated in Fig. 1, and determines to which side of the boundary the value of the dimension falls (Ashby & Townsend, 1986). The perception of stimulus values along each dimension is assumed to be noisy; hence, there is a normal distribution of percepts (see Fig. 3) associated with each stimulus value along each dimension. The total probability of taking a step toward the $+A$ boundary is determined by the proportion of the distribution that falls in the A region of the stimulus space:

$$p = \int_{Di}^{\infty} N\left(x_i, \mu_D, \sigma_D^2\right) dx_i, \tag{5}$$

where D_i is the boundary position on dimension D, and the percept on each dimension is normally distributed with mean μ_D and variance σ^2_D. For coactive models, the single-channel random walk is driven by the joint bivariate normal distribution on x and y (see, e.g., Little et al., 2012), with mean vector μ_{xy} and covariance $\sum_{xy}$; hence,

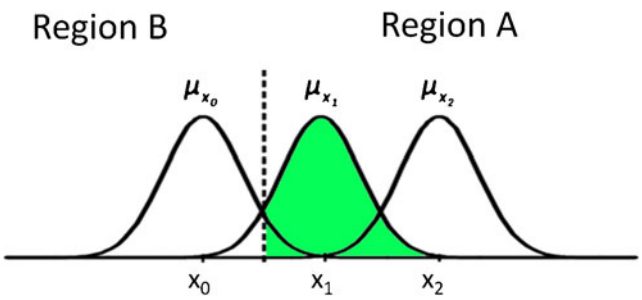

Fig. 3 Percepts sampled from distribution x_1 that fall on the region-A side of the decision boundary lead the random walk on dimension x to take steps toward criterion $+A$

$$p = \int_{D_y}^{\infty} \int_{D_x}^{\infty} MVN\left(x_i, y_j, \mu_{xy}, \sum_{xy}\right) dx_i dx_j. \tag{6}$$

In general, stimuli with values that lie far from the decision bound (e.g., value x_2 in Fig. 1) will lead to faster and more accurate decisions than will those that lie close (e.g., value x_1 in Fig. 1), because the random walk will take more consistent steps toward the appropriate response criterion.

Note that for the serial and parallel models, described in the following sections, the probabilities for taking a step toward the $+A$ boundary are determined separately for each of the random walks in the model, using Eq. 5 with a mean, μ_{Dx}, and variance, σ^2_{Dx}, for dimension x and a separate mean, μ_{Dy}, and variance, σ^2_{Dy} for dimension y. In Fifić et al. (2010) and Little et al. (2011), the stimulus dimensions were assumed to be perceptually independent; that is, $\mu_{Dx} \neq \mu_{Dy}$ *and* $s^2_{Dx} \neq s^2_{Dy}$.[3] Furthermore, the two channels were assumed to be decisionally independent; that is, the boundaries on each dimension, D_x and D_y, were assumed to be orthogonal to the dimensional axis. A further strength of the parametric approach is that these assumptions can be systematically varied to examine the resulting RT predictions *independent of any assumptions about the underlying architecture*. This exploration is beyond the scope of the present article but is addressed elsewhere (Little et al., 2012).

Summary

Within the logical-rule modeling framework, single-channel random-walk models are used to model the decision times associated with individual stimulus dimensions. Several existing models of RTs also employ single-channel architectures (e.g., EBRW, Nosofsky & Palmeri, 1997; stochastic GRT, Ashby, 2000; and the integrated theory of attention, Logan, 2002). Single-channel coactive models are also used within the logical-rule framework to capture performance on stimuli that require pooling across stimulus dimensions prior to initiating the decision process (e.g., integral stimuli such as colors, which comprise varying values of brightness and saturation; Little et al., 2012). Other types of decisions that involve combining independent decisions about stimulus dimensions (e.g., separable-dimensioned stimuli; Little et al., 2011) or, to use a related example, searching through multiple elements in a visual search array (Thornton & Gilden, 2007) require the combination of separate random-walk processes either in serial or in parallel. These architectures are outlined in the following sections.

[3] The variance parameters were freely estimated, so it was possible but unlikely that σ^2_{Dx} might equal σ^2_{Dy}.The means of the perceptual distributions were set equal to values determined via a multidimensional-scaling solution to similarity comparison data.

Serial models

In order to generate predictions for a serial processing model, we need to take into account the order in which the multiple single-channel random-walk processes occur. That is, to compute the expected decision times and distributions of decision times for the serial model, we need to take into account the probabilities that process x is undertaken before process y and that process y is undertaken before process x. Conceptually, for a serial self-terminating model, if both of the processes end at the $-B$ boundary, then the response is B and the decision time is given by whichever process occurred first. If only one process ends at the $-B$ boundary, then the response is B, and the decision time is given by the elapsed time of that process IF that process occurred first; if the B-terminating process occurred second, then the decision time is given by the sum of the elapsed times of both processes. Likewise, if both of the processes end at the $+A$ boundary, then the decision time is given by the sum of the elapsed times of both processes, and the predicted response is A. The notation adopted to describe the serial model is shown in Table 1.

In this section, we derive predictions for two processes, process x and process y, but extension to more than two processes would be straightforward (so long as one accounts for the combinatorial probability that each process occurs before and after every other process).

State space of the serial model

With two processes, there are four potential outcome states: (1) each process terminates at the $+A$ boundary (the upper boundary), $P(R_x = A, R_y = A)$; (2) the first process terminates at the $+A$ boundary and the second process terminates that the $-B$ boundary, $P(R_x = A, R_y = B)$; (3) the first process terminates at the $-B$ boundary and the second process terminates at the $+A$ boundary, $P(R_x = B, R_y = A)$; and (4) both processes terminate at the $-B$ boundary (the lower boundary), $P(R_x = B, R_y = B)$. Because the processes are independent, the probability of each of the states is given as follows:

Table 1 Mathematical notation used to describe the serial processing model

Symbol	Meaning
A	Category A decision
B	Category B decision
$E_x[N]$, $E_y[N]$	Expected number of steps for process x and process y, respectively
f_x, f_y	Distribution of step probabilities for process x and process y, respectively
R_i	Response for process $i \in \{x, y\}$
R	Overall response
X1	Process x goes first
Y1	Process y goes first

$$
\begin{aligned}
P(R_x = A, R_y = B) &= P(R_x = A) \times \left[1 - P(R_y = A)\right],\\
P(R_x = B, R_y = A) &= \left[1 - P(R_x = A)\right] \times P(R_y = A),\\
P(R_x = B, R_y = B) &= \left[1 - P(R_x = A)\right] \times \left[1 - P(R_y = A)\right].
\end{aligned}
\tag{7}
$$

Choice probabilities for the serial model

If we assume that the model predicts Category B any time that either of the single-channel processes terminates at $-B$, in accord with Fig. 1, then the overall probability that the model predicts a B response is $P(R = B) = P(R_x = B, R_y = B) + P(R_x = A, R_y = B) + P(R_x = B, R_y = A)$. The overall probability that the model predicts an A response is $P(R = A) = P(R_x = A, R_y = A) = 1 - P(R = B)$.

Expected decision time for the serial model

Note the following:

$$
P(X1) = P(X1, Y2) = P(Y2) = 1 - P(X2). \tag{8}
$$

Formally, the overall expected decision time for the serial model is a weighted average of the relevant outcomes, given as follows:

$$
\begin{aligned}
E[N] = {} & P(R_x = B,\ R_y = B)\left[P(X1)E_x[N|R_x = B] + (1 - P(X1))E_y\left[N|R_y = B\right]\right]\\
& + P(R_x = A, R_y = B)\left[P(X1)\left(E_x[N|R_x = A] + E_y\left[N|R_y = B\right]\right) + (1 - P(X1))E_y\left[N|R_y = B\right]\right]\\
& + P(R_x = B, R_y = A)\left[P(X1)E_x[N|R_x = B] + (1 - P(X1))\left(E_x[N|R_x = B] + E_y\left[N|R_y = A\right]\right)\right]\\
& + P(R_x = A, R_y = A)\left(E_x[N|R_x = A] + E_y\left[N|R_y = A\right]\right).
\end{aligned}
\tag{9}
$$

Here, $E_x[N|\ R_x = A]$ refers to the expected duration of process x whenever x terminates at the $+A$ boundary, and $E_y[N|\ R_y = A]$ refers to the expected duration of process y whenever y terminates at the $+A$ boundary (see Eq. 2). In applications of the mixed-order serial model (Fifić et al., 2010; Little et al., 2011), the sequence probability $P(X1)$ is estimated as a free parameter.

Distribution of decision times for the serial model

The computation of the probabilities that a serial process ends after the Nth step is very similar to the calculation used for the expected decision time in Eq. 9; however, the expected outcomes of processes x and y are replaced with the decision time distributions computed in Eq. 3. When at

least one of the processes ends at $-B$, the distribution of decision times is given as

$$\begin{aligned} f[N, R=B] = {} & P(R_x=B, R_y=B)\ \left[P(X1) f_x[N|R_x=B] + (1-P(X1)) f_y\left[N|R_y=B\right]\right] \\ & + P(R_x=A, R_y=B)\ \left[P(X1)\left(\left[f_x * f_y\right]\left(N|R_x=A, R_y=B\right)\right) + (1-P(X1)) f_y\left[N|R_y=B\right]\right] \\ & + P(R_x=B, R_y=A)\ \left[P(X1) f_x[N|R_x=B] + (1-P(X1))\left(\left[f_x * f_y\right]\left(N|R_x=B, R_y=A\right)\right)\right], \end{aligned} \tag{10}$$

where $\left[f_x * f_y\right]\left(N|R_x=A, R_y=B\right)$ is the convolution of the first-passage time distributions from each individual process. When both processes end at the $+A$ boundary, the distribution of decision times is

$$\begin{aligned} f[N, R=A] = {} & P(R_x=A, R_y=A) \\ & \times \left[f_x * f_y\right]\left(N|R_x=A, R_y=A\right) \end{aligned} \tag{11}$$

Equations 10 and 11 are for a self-terminating serial model. The equation for an exhaustive model (e.g., a model that is serial exhaustive, regardless of the boundary termination of any single random walk) is simply $[f_x * f_y]$. Once the RT distribution of the serial process is determined, the nondecision time distribution can be convolved as shown in Eq. 4.

Summary

Within the logical-rule modeling framework, serial models are used to model the decision times associated with stimulus dimensions that are processed sequentially. For instance, when visual features are spatially separated, participants learn to process these features in a serial order (Fifić et al., 2010; Little et al., 2011). By contrast, Little et al. (2011) found that when visual features spatially overlapped, processing was better described as a mixture of serial and parallel processing. In the next section, the computations for the parallel model are described.

Parallel models

Parallel logical-rule models, like the serial logical-rule models, require a combination of multiple independent processes; however, in a parallel model, the individual walks occur simultaneously rather than sequentially. Hence, different assumptions are needed in order to compute the relevant predictions.

State space of the parallel model

The four outcome states for the parallel model are identical to the outcome states of the serial model [e.g., $P \times \left(R_x=A, R_y=A\right), P\left(R_x=A, R_y=B\right), P\left(R_x=B, R_y=A\right)$, and $P\left(R_x=B, R_y=B\right)$]. Hence, the overall probability that the model predicts an A or a B response is the same as for the serial model.

Expected decision time for the parallel model

For a parallel self-terminating model, if both of the processes end at the $-B$ boundary, then the response is B, and the decision time is given by whichever process terminated first (i.e., the minimum processing time). If only one of the parallel processes ends at the $-B$ boundary, then the response is B and the decision time is given by the elapsed time of only that process. If both of the processes end at the $+A$ boundary, then the decision time is determined by whichever process took *longer* to terminate (i.e., the maximum processing time).

The overall expected decision time for the model is a weighted average of the relevant outcomes, given as follows:

$$\begin{aligned} E(N) = {} & P\left(R_x=B, R_y=B\right) E[f_{\min}(N)] \\ & + P\left(R_x=A, R_y=B\right) E_y\left[N|R_y=B\right] \\ & + P\left(R_x=B, R_y=A\right) E_x[N|R_x=B] \\ & + P\left(R_x=A, R_y=A\right) E[f_{\max}(N)], \end{aligned} \tag{12}$$

where $E[f_{\min}(N)]$ is the expected number of steps from the distribution of minimum step times for process x and process y, and $E[f_{\max}(N)]$ is the expected number of steps from the distribution of maximum step times for process x and process y.

Deriving RT distribution predictions for the parallel model requires the computation of the distribution of minimum and maximum finishing times for the independent random walks. As discussed above, the expected number of steps is given by the expected value of the distribution of maximum termination times. The cumulative distribution of the maximum termination times is computed as the product of the cumulative distribution, $F(N) = P[n <= N] = \sum_{i=1}^{N} f(i)$, over steps for each of the two processes (see, e.g., Ratcliff, 1978).

The cumulative distribution for the maximum process time and minimum process time are

$$F_{\max}(N) = \prod_{i=1}^{N\ processes} F_i(N), \tag{13}$$

$$F_{\min}(N) = 1 - \prod_{i=1}^{N\ processes} 1 - F_i(N). \quad (14)$$

Hence, the probability mass function of the maximum processing times is

$$f_{\max}(N) = \Delta_n F_{\max}(N), \quad (15)$$

and the probability mass function of the minimum processing times is

$$f_{\min}(N) = \Delta_n F_{\min}(N) \quad (16)$$

(see, e.g., Luce, 1986, p. 9). The Δ value is the discrete derivative, which is computed as the difference between the cumulative probabilities at step $N+1$ minus the cumulative probabilities at step N (where the probability of $N_0=0$).

Distribution of decision times for the parallel model

As with the serial model, to compute the distribution of decision times for the parallel model we replace the expectations in Eq. 12 with the decision time distributions from each component random walk. If at least one process ends at the $-B$ boundary, the decision time probabilities are

$$\begin{aligned} f[N, R=B] =& P(R_x = B, R_y = B) f_{\min}(N) \\ &+ P(R_x = A, R_y = B) f_y[N|R_y = B] \\ &+ P(R_x = B, R_y = A) f_x[N|R_x = B] \end{aligned} \quad (17)$$

If both processes end at the $+A$ boundary, the decision time probabilities are

$$f[N, R=A] = P(R_x = A, R_y = A) f_{\max}(N). \quad (18)$$

Summary

Parallel models are used to model the decision times associated with stimulus dimensions that are processed simultaneously. For the two-dimensional categorization task shown in Fig. 1, parallel processing has only been demonstrated for visually overlapping stimulus features. Furthermore, this parallel processing was part of a trial-by-trial mixture of serial and parallel processing (Little et al., 2011). The parametric instantiation of the serial and parallel models presented here can easily be combined by assuming that the processing is serial on some proportion of the trials and by taking a weighted mixture of the decision-time distributions. In other domains, parallel processing is common; for instance, parallel processes are often implicated in fast, "pop-out" visual search tasks (Thornton & Gilden, 2007). However, parallel processes may not always be of unlimited capacity, as presented here. Instead, interactive parallel models may have processing rates that are affected by the total number of random walks running concurrently (Townsend, 1990). The logical-rule models could be extended to model interactive parallel processing by allowing step counts to be shared between channels (cf. Eidels et al., 2011).

Discussion

Theoretical and methodological advances have made it possible to design experiments that can differentiate serial and parallel models (i.e., the systems factorial technology; Fifić, Nosofsky, & Townsend, 2008; Townsend & Nozawa, 1995; Townsend & Fifić, 2004; Townsend & Wenger, 2004a). Even research in visual search, where the ambiguity of the existing data has been enough to prompt the declaration of the serial-versus-parallel debate as a dead-end, has shown renewed interest in differentiating serial and parallel models by using better experimental techniques (Thornton & Gilden, 2007) and by focusing on full RT distributions (Donkin & Shiffrin, 2011; Wolfe, Palmer, & Horowitz, 2010). In categorization, questions about the serial versus parallel processing of stimuli comprising multiple dimensions have only recently begun to be asked (Bradmetz & Mathy, 2008; Fifić et al., 2010; Fifić et al., 2008; Lafond, Lacouture, & Mineau, 2007; Little et al., 2011). Consequently, having models that can produce predictions at the level of full RT distributions is central to uncovering the underlying architecture used to process multidimensional stimuli. The computations provided in this report allow for efficient predictions to be generated from all of the logical-rule categorization models.

The focus of this report has been on discrete-time random-walk models; however, any sequential-sampling model could be substituted (e.g., continuous-time diffusion [Ratcliff, 1978] or accumulator models [Brown & Heathcote, 2009]). Furthermore, any distribution over RTs can be substituted for the random-walk model. For example, Townsend and Ashby (1983) used an exponential distribution to model RTs from a variety of serial and parallel processes. Likewise, other methods may be used to generate the step probabilities, p and q; for example, EBRW (Nosofsky & Palmeri, 1997) uses the summed similarity between a probe stimulus and stored exemplars to compute p; Smith and Ratcliff (2009) use a model of visual short-term memory to drive the decision-making process; and both Ratcliff, Van Zandt, and McKoon (1999) and Donkin, Brown, Heathcote, and Wagenmakers (2011) freely estimate the rate at which evidence is accumulated toward each boundary.

Recently, nonparametric methods have been developed to differentiate more complex architectures, such as interactive parallel models, with facilitative or inhibitory cross-talk between the processing channels (Eidels et al., 2011). In addition, nonparametric measures of capacity and parallel interaction have been shown to correspond closely to the drift rate in parametric sequential-sampling models such as

the single-channel models described here (Eidels et al., 2010). Future work will be needed to derive predictions for more complicated parametric parallel models in order to capture the processes underlying facilitatory and inhibitory interactions in categorization.

Author note Preparation of this article was facilitated by ARC Discovery Grant DP120103120. MATLAB code for the models described here, along with a description of a parameter estimation routine, are available in the supplementary materials. The software is available at www.psych.unimelb.edu.au/research/labs/knowlab/documents/LogicalRuleModels.zip.

References

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology, 44,* 210–329.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179. doi:10.1037/0033-295X.93.2.154

Bourne, L. E. (1970). Knowing and using concepts. *Psychological Review, 77,* 546–556.

Bradmetz, J., & Mathy, F. (2008). Response times seen as decompression times in Boolean concept use. *Psychological Research, 72,* 211–234.

Brown, S., & Heathcote, A. (2009). A ballistic model of choice response time. *Psychological Review, 112,* 117–128.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York, NY: Wiley.

Busemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 538–564.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. New York, NY: Sage.

Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology, 47,* 304–322.

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review, 18,* 61–69. doi:10.3758/s13423-010-0022-4

Donkin, C., & Shiffrin, R. M. (2011). Visual search as a combination of automatic and attentive processes. In L. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *Expanding the space of cognitive science: Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 2830–2835). Austin, TX: Cognitive Science Society.

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

Eidels, A., Houpt, J. W., Altieri, N., Pei, L., & Townsend, J. T. (2011). Nice guys finish fast and bad guys finish last: Facilitatory vs. inhibitory interaction in parallel systems. *Journal of Mathematical Psychology, 55,* 176–190.

Feller, W. (1968). *An introduction to probability theory and its applications*. New York, NY: Wiley.

Fifić, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review, 117,* 309–348. doi:10.1037/a0018526

Fifić, M., Nosofsky, R. M., & Townsend, J. T. (2008). Information-processing architectures in multidimensional classification: A validation test of the systems factorial technology. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 356–375. doi:10.1037/0096-1523.34.2.356

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

Lafond, D., Lacouture, Y., & Mineau, G. (2007). Complexity minimization in rule-based category learning: Revising the catalog of Boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology, 51,* 57–74.

Levine, M. (1975). *A cognitive theory of learning: Research on hypothesis testing*. Hillsdale, NJ: Erlbaum.

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

Little, D. R., Nosofsky, R. M., Donkin, C., & Denton, S. E. (2012). *Logical rules and the classification of integral-dimension stimuli.* Manuscript submitted for publication.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review, 109,* 376–400.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York, NY: Oxford University Press.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14,* 247–279. doi:10.1016/0010-0285(82)90010-X

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300. doi:10.1037/0033-295X.104.2.266

Palmer, J., & McLean, J. (1995, August). *Imperfect, unlimited capacity, parallel search yields large set-size effects*. Paper presented at the meeting of the Society for Mathematical Psychology, Irvine, CA.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108. doi:10.1037/0033-295X.85.2.59

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106,* 261–300. doi:10.1037/0033-295X.106.2.261

Schweickert, R. (1992). A silver jubilee. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 535–566). Cambridge, MA: MIT Press.

Smith, P. L. (2010). From Poisson shot noise to the integrated Ornstein–Uhlenbeck process: Neurally principled models of information accumulation in decision-making and response time. *Journal of Mathematical Psychology, 54,* 266–283.

Smith, P. L., & McKenzie, C. R. L. (2012). Diffusive information accumulation by minimal recurrent neural models of decision making. *Neural Computation*.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27,* 161–168.

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review, 116,* 282–317.

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

Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review, 114,* 71–103.

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

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1,* 46–54. doi:10.1111/j.1467-9280.1990.tb00067.x

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

Townsend, J. T., & Fifić, M. (2004). Parallel versus serial processing and individual differences in high-speed search in human memory. *Perception & Psychophysics, 66,* 953–962. doi:10.3758/BF03194987

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39,* 321–359. doi:org/10.1006/jmps.1995.1033

Townsend, J. T., & Wenger, M. J. (2004a). The serial–parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review, 11,* 391–418. doi:10.3758/BF03196588

Townsend, J. T., & Wenger, M. J. (2004b). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review, 111,* 1003–1035. doi:10.1037/0033-295X.111.4.1003

Trabasso, T., & Bower, G. H. (1968). *Attention in learning: Theory and research*. New York, NY: Wiley.

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion Model Analysis with MATLAB: A DMAT Primer. *Behavior Research Methods, 40*, 61–72.

Voss, A., & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology, 52,* 1–9. doi:10.1016/j.jmp.2007.09.005

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14,* 3–22. doi:10.3758/BF03194023

Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research, 50,* 1304–1311.