



# Sequential sampling model for multiattribute choice alternatives with random attention time and processing order

Adele Diederich<sup>1\*</sup> and Peter Oswald<sup>2</sup>

<sup>1</sup> Cognitive Psychology, School of Humanities and Social Sciences, Jacobs University, Bremen, Germany

<sup>2</sup> Mathematics, Modeling, and Computing Center, School of Engineering and Science, Jacobs University, Bremen, Germany

## Edited by:

José Antonio Díaz, Universidad de Granada, Spain

## Reviewed by:

Chris Donkin, University of New South Wales, Australia

José Antonio Díaz, Universidad de Granada, Spain

## \*Correspondence:

Adele Diederich, Cognitive Psychology, School of Humanities and Social Sciences, Jacobs University, Campus Ring 1, Bremen 28759, Germany  
e-mail: a.diederich@jacobs-university.de

A sequential sampling model for multiattribute binary choice options, called *multiattribute attention switching* (MAAS) model, assumes a separate sampling process for each attribute. During the deliberation process attention switches from one attribute consideration to the next. The order in which attributes are considered as well for how long each attribute is considered—the attention time—influences the predicted choice probabilities and choice response times. Several probability distributions for the attention time with different variances are investigated. Depending on the time and order schedule the model predicts a rich choice probability/choice response time pattern including preference reversals and fast errors. Furthermore, the difference between finite and infinite decision horizons for the attribute considered last is investigated. For the former case the model predicts a probability  $p_0 > 0$  of not deciding within the available time. The underlying stochastic process for each attribute is an Ornstein-Uhlenbeck process approximated by a discrete birth-death process. All predictions are also true for the widely applied Wiener process.

**Keywords:** sequential sampling, multiattribute, attention time, time schedule, order schedule, finite time horizon, Ornstein-Uhlenbeck, Wiener

## 1. INTRODUCTION

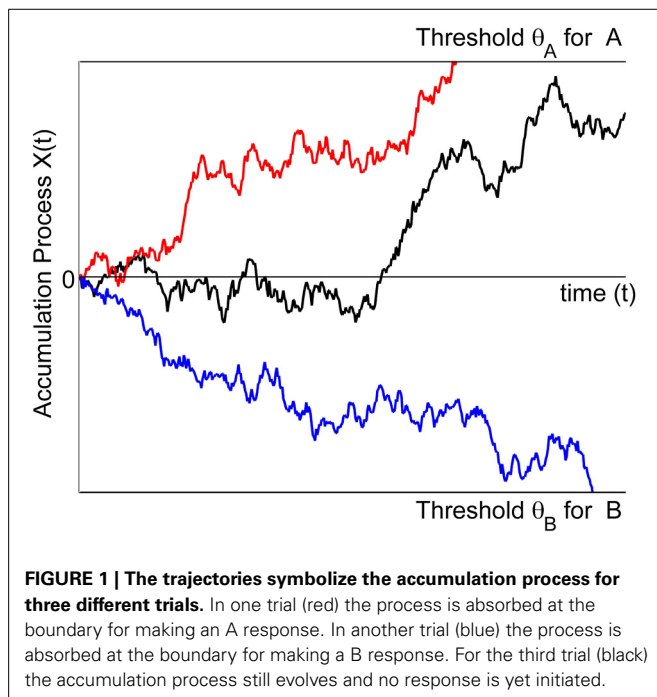
Sequential sampling models are powerful models to account simultaneously for choice probabilities and choice response times. They have become the dominant approach to modeling decision processes in cognitive science. Their application includes a variety of psychological tasks from basic perceptual decision to complex preferential choice tasks. Early on they have been applied to identification and discrimination tasks (e.g., Edwards, 1965; Laming, 1968; Pike, 1973; Link and Heath, 1975; Heath, 1981; Ashby, 1983); memory retrieval (e.g., Stone, 1960; Ratcliff, 1978; Van Zandt et al., 2000); and classification (e.g., general recognition theory, Ashby, 2000; exemplar-based random walk models of classification, Nosofsky and Palmeri, 1997) to account for speed-accuracy data. They have also been used for preferential decision tasks (e.g., decision field theory (DFT), Busemeyer and Townsend, 1993; multiattribute dynamic decision model, Diederich, 1997; Diederich and Busemeyer, 1999) to account for choice response times and choice probabilities interpreted as preference strength; judgment and confidence ratings (Pleskac and Busemeyer, 2010); to account for selling prices, certainty equivalents, and preference reversal phenomena (Busemeyer and Goldstein, 1992; Johnson and Busemeyer, 2005). More recently, they have been applied to combining perceptual decision making and payoffs (Diederich and Busemeyer, 2006; Diederich, 2008; Rorie et al., 2010; Gao et al., 2011). Furthermore, these models have been closely linked to measures from neuroscience like multi-cell electrode recordings (e.g., Ditterich, 2006; Gold and Shadlen, 2007; Churchland et al., 2008).

Sequential sampling models assume that (1) stimulus and choice alternative characteristics can be mapped onto a hypothetical numerical value representing the instantaneous level of evidence (activation, information, or preference—the wording often depends on the context), (2) some random fluctuation of this value over time occurs, (3) this evidence is accumulated over time, and (4) a final choice is made as soon as the evidence reaches a threshold. Therefore, sequential sampling can be described as a stochastic process. Two quantities are of foremost interest: (1) the probability that the process eventually reaches one of the thresholds or boundaries for the first time (the criterion to initiate a response), i.e., *first passage probability*; (2) the time it takes for the process to reach one of the boundaries for the first time, i.e., *first passage time*. The former quantity is related to the observed relative frequencies, the latter usually to the observed mean choice response times or the observed choice response time distribution.

Two classes of sequential sampling models have been predominantly used in psychology: Random walk/diffusion models and accumulator/counter models. The former are typically applied to a binary choice task, so that evidence for one choice alternative is at the same time evidence against the other. A decision is made as soon as the process reaches one of two preset criteria. In the latter, an accumulator/counter is established for each choice alternative separately, and evidence is accumulated in parallel. A decision is made as soon as one counter wins the race to reach one preset criterion. The accumulators/counters may or may not be independent. In the following we focus on random

walk/diffusion models. For a review of both diffusion models and counter models see Ratcliff and Smith (2004).

To be more precise and to introduce notation, let  $X(t)$  denote the accumulation process. For a binary choice, say between choice options A and B (Figure 1), the models assume that the decision process begins with an initial state of evidence  $X(0)$ . This initial state may either favor option A ( $X(0) > 0$ ) or option B ( $X(0) < 0$ ) or may be neutral with respect to A or B ( $X(0) = 0$ ). Upon presentation of the choice options, the decision maker sequentially samples information from the stimulus display over time, retrieves information from memory, or forms preferences, depending on the context. The small increments of evidence sampled at any moment in time are such that they either favor option A ( $dX(t) > 0$ ) or option B ( $dX(t) < 0$ ). The evidence is accumulated from one moment in time to the next by summing the current state with the new increment:  $X(t+h) \approx X(t) + \mu(X(t), t)h + \sigma(X(t), t)(W(t+h) - W(t))$ . Here,  $\mu(x, t)$  is called the *drift rate* and describes the expected value of increments per unit time. The factor  $\sigma(x, t)$  in front of the increments  $W(t+h) - W(t)$  of a standard Wiener process  $W(t)$  is called the *diffusion rate*, and relates to the variance of the increments. This process continues until the magnitude of the cumulative evidence exceeds a threshold criterion,  $\theta$ . The process stops and option A is chosen as soon as the accumulated evidence reaches a criterion value for choosing A (here,  $X(t) = \theta_A > 0$ ) or it stops and chooses option B as soon as the accumulated evidence reaches a criterion value for choosing B (here  $X(t) = \theta_B < 0$ ). The probability of choosing A over B is determined by the accumulation process reaching the threshold for A before reaching the threshold for B. The criterion is assumed to be set by the decision maker prior to the decision task.



The Wiener process with drift, lately called *drift-diffusion model* in the psychological literature (Bogacz et al., 2006), is the most widely applied model. Different versions reflect additional assumptions for specific psychological domains. Ratcliff (1978) proposed a diffusion model for memory retrieval that is used for various psychological decision tasks. It is based on the work by Laming (1968) and Link and Heath (1975) and assumes variability in the starting point (i.e.,  $X(0)$  follows a uniform distribution), and the drift rate  $\mu = \mu(t)$  of the Wiener process is normally distributed (cf. Laming). The residual time, i.e., the time other than the decision time, such as stimulus encoding and motor response, is assumed to be uniformly distributed and added to the decision time, i.e., response time equals the decision time plus a residual (non-decision) time. For a recent overview with applications see Voss et al. (2013). Other approaches include the Ornstein-Uhlenbeck model that linearly accumulates evidence with decay (Busemeyer and Townsend, 1993; Diederich, 1997), and the leaky competing accumulator model (Usher and McClelland, 2001) that non-linearly accumulates evidence with decay.

Common to almost all of these approaches is the assumption that a single integrated source of evidence generates the evidence during the deliberation process leading to a decision. In particular, the integrated source may be based on multiple features or attributes, but all of these features or attributes are assumed to be combined and integrated into a single source of evidence, and this single source is used throughout the decision process until a final decision is reached. Diederich (e.g., Diederich, 1995, 1997, 2003, 2008), however, assumed a separate process for each attribute<sup>1</sup>. The decision maker switches attention from one attribute to the next during the time course of one trial. For instance, in a crossmodal task (visual, auditory, tactile), Diederich (1995) assumed a serial processing controlled by stimulus input at given stimulus onset asynchronies (SOA). That is, the order of attributes, here a light, followed by a tone, followed by a tactile vibration, as well as the point in time when a new attribute was added, here the tone presented at  $t_1$  ( $t_1$  ms after the light onset) and the tactile vibration at  $t_2$  ( $t_2$  ms after the light onset) was determined externally by the experimental setup. In the following we will call an attention switch at predetermined, fixed times, and predefined order attributes, a *deterministic time and order schedule*. Often, however, neither the processing order of attributes nor the point in time when the decision maker switches attention from one attribute to the next are known or can be inferred from the experimental setup. For those cases, Diederich (1997) proposed a specific model in which attention switches from one attribute to the next with some probability. This is an instance of a *random time and order schedule* which will be investigated more systematically in the present study.

<sup>1</sup>The notion of *attributes* is defined here in a broad sense. For example, it includes dimensions such as color and size of visual target; amplitude and frequency of a tone; different modalities in a crossmodal task; payoff information and perceptual information; attitudinal evidence and perceptual evidence; prize and quality of a consumer product and more.

The purpose of this paper is to present a unified treatment of sequential sampling models for both deterministic and random time and order schedules. To do so we start with deriving expressions for mean choice response times and choice probabilities for a deterministic time and order schedule before we show how they extend to random time and order schedules, including Poisson, binomial, geometric, and uniform distributions for the attention time devoted to each attribute in the sequence before attention switches to the next randomly or deterministically chosen attribute. We will provide first numerical evidence on the influence of various properties of a schedule on the predictions for mean choice response times and choice probabilities.

## 2. PRELIMINARIES

The model applies to any finite number of attributes that the decision maker may consider, i.e.,  $k = 1, \dots, K$ . For convenience we first describe the process for one attribute. As underlying information process for each attribute we assume an Ornstein-Uhlenbeck process  $X(t)$  defined by

$$dX(t) = (\delta_k - \gamma_k X(t)) dt + \sigma_k dW(t), \tag{1}$$

where  $W(t)$  is a standard Wiener process. The parameters  $\delta_k$ ,  $\gamma_k$ , and  $\sigma_k$  are characteristics of the  $k$ -th attribute. The attribute characteristics may affect the quality of the extracted evidence for choosing  $A$  over  $B$  and this quality of evidence determines the drift rate  $\delta_k$ . That is, the better an attribute discriminates between  $A$  and  $B$ , the larger is  $\delta_k$ . The parameter  $\gamma_k$  which induces a change of the drift rate depending on the current state in the state space is often connected to memory processes (e.g., primacy and recency effects), conflict situations (e.g., approach-avoidance), or similarities between choice alternatives. Thus, together the effective drift  $\delta_k - \gamma_k X(t)$  determines the direction and the velocity of the process when considering the  $k$ -th attribute at time  $t$ . Note that by setting  $\gamma_k$  to 0 results in a Wiener process with drift. That is, all the analysis we perform in the following is also valid for the Wiener process with drift. The diffusion coefficient  $\sigma_k$  indicates the variance of the increments of the process, for simplicity, we will set  $\sigma_k = \sigma$  for all  $k$ .

### 2.1. MATRIX APPROACH

Stochastic processes such as the above  $X(t)$  can be approximated by a discrete time, finite state space Markov chain. We use the matrix approach since it is simple to implement, sufficient in determining the entities of interest, i.e., choice probabilities and choice response times, and flexible to account for non-stationary and non-linear properties one wishes to include for the decision making process in the future. The continuous state space  $[\theta_B, \theta_A]$  of the piecewise Ornstein-Uhlenbeck process  $X(t)$  is replaced by a finite state space  $S = \{-m_B, \dots, m_A\}$  with  $m = m_A + m_B + 1$  states. The diffusion process  $\{X(t), t \geq 0\}$  is approximated by a discrete random walk  $\{\tilde{X}(n), n \geq 0\}$  with values in  $S$  such that  $X(n\tau) \approx \Delta \cdot \tilde{X}(n)$  and  $\theta_A \approx m_A \Delta$  and  $\theta_B \approx -m_B \Delta$ , where  $\Delta$  is the step size of change in evidence. To achieve convergence in the limit, the discretization parameters ( $\Delta$  for state space, and  $\tau$  for time) are tied to each other by the relation  $\Delta = \sigma \sqrt{\tau}$ .

The attribute-related matrices  $P_k$ ,  $k = 1, \dots, K$ , are given in their canonical form by

$$P_k = \left[ \begin{array}{c|c} I & 0 \\ \hline R_k & Q_k \end{array} \right]$$

$$= \left[ \begin{array}{cc|cccc} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ \hline p_{21}^{(k)} & 0 & p_{22}^{(k)} & p_{23}^{(k)} & \dots & 0 & 0 \\ 0 & 0 & p_{32}^{(k)} & p_{33}^{(k)} & \dots & 0 & 0 \\ 0 & 0 & 0 & p_{43}^{(k)} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & p_{m-3,m-2}^{(k)} & 0 \\ 0 & 0 & 0 & 0 & \dots & p_{m-2,m-2}^{(k)} & p_{m-2,m-1}^{(k)} \\ 0 & p_{m-1,m}^{(k)} & 0 & 0 & \dots & p_{m-1,m-2}^{(k)} & p_{m-1,m-1}^{(k)} \end{array} \right] \tag{2}$$

where

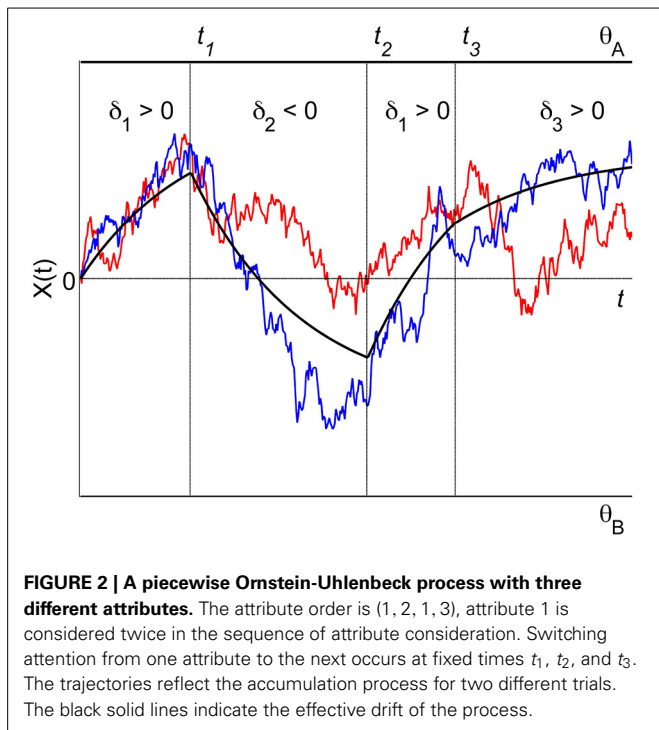
$$p_{i,i-1}^{(k)} = \frac{1}{2} \left( 1 - (\delta_k - \gamma_k i \Delta) \frac{\sqrt{\tau}}{\sigma} \right),$$

$$p_{i,i+1}^{(k)} = \frac{1}{2} \left( 1 + (\delta_k - \gamma_k i \Delta) \frac{\sqrt{\tau}}{\sigma} \right),$$

for  $i = 2, \dots, m - 1$  (here, the index  $i$  corresponds to the state  $i - 1 - m_B$ ). As  $\Delta \rightarrow 0$  (or, equivalently,  $\tau \rightarrow 0$ ), the decision probabilities and mean choice response times obtained from the Markov chain model converge to the values obtained from the underlying continuous process  $X(t)$ . The identity matrix  $I$  corresponds to the two absorbing states ( $-m_B$  and  $m_A$ ) associated with the two decision thresholds, one for each choice alternative; the matrix  $Q_k$  contains the transient probabilities, corresponding to the updating evidence process, and the matrix  $R_k$  contains the one-step transition probabilities from the transient to the absorbing states. In particular, the first column vector of the matrix  $R_k$  (denoted by  $R_{B,k}$ ) contains the transient probabilities for reaching alternative  $B$ , while the second  $R_{A,k}$  contains the ones for alternative  $A$ . For details and derivations see Diederich (1997) and Diederich and Busemeyer (2003).

### 2.2. TIME AND ORDER SCHEDULE

For  $K$  attributes, each one to be considered for some specific time in some specific order it is convenient to introduce a formal schedule of both time and order. A finite time and order schedule consists of a set of  $L$  consecutive time intervals  $\{[t_{l-1}, t_l]\}_{l=1,\dots,L}$  and the attribute sequence  $\{k_l \in \{1, \dots, K\}\}_{l=1,\dots,L}$  which specifies that during the time interval  $[t_{l-1}, t_l]$  the  $k_l$ -th attribute is considered. At switching time  $t_l$ ,  $l = 1, \dots, L - 1$ , attention switches from attribute  $k_l$  to attribute  $k_{l+1}$ . Depending on the situation, the final time  $t_L$  may be set finite (then the decision process may also finish without deciding for one of the alternatives) or infinite. Consequently, the process  $X(t)$  determined by such a schedule is a piecewise Ornstein-Uhlenbeck process, defined over a finite partition  $t_0 = 0 < t_1 < \dots < t_{L-1} < t_L \leq +\infty$  of the time interval  $[0, t_L]$ , where for  $t \in [t_{l-1}, t_l]$  the process is determined by (1) with  $k = k_l$ . **Figure 2** shows an example with three



different attributes ( $K = 3$ ) and a deterministic time and order schedule of length  $L = 4$  with switching times  $t_l$  independent of the trajectories, and attribute order (1, 2, 1, 3), i.e.,  $k_1 = 1, k_2 = 2, k_3 = 1, k_4 = 3$  (note that the first attribute is reconsidered once).

For fixed  $\Delta$  resp.  $\tau$ , the  $m \times m$  transition probability matrix  $\tilde{P}_n$  containing the transition probabilities  $\tilde{P}_{ij} := P(\tilde{X}_{n+1} = i | \tilde{X}_n = j)$  for the  $n$ -th step of the discrete-time random walk depends on the currently considered attribute defined by the time and order schedule, i.e., we set  $\tilde{P}_n = P_{k_l}$  if  $n = n_{l-1}, \dots, n_l - 1$ , where  $n_0 = 0, \tau n_l \approx t_l$  for  $l = 1, \dots, L$  (if  $t_L = \infty$ , we formally set  $n_L = \infty$ ).

### 3. CHOICE PROBABILITIES AND MEAN CHOICE RESPONSE TIMES

In this section we derive the choice probabilities and mean choice response times for various time and order schedules. For simplicity we assume an unbiased process, i.e., with  $X(0) = 0$  and symmetric decision thresholds, i.e.,  $\theta_A = -\theta_B$ . Since the diffusion coefficient is a scaling parameter it will be set to  $\sigma = 1$  for all attributes throughout. We start with the deterministic time and order schedule.

#### 3.1. DETERMINISTIC TIME AND ORDER SCHEDULE

The evidence accumulation process for attribute  $k_1$ , which is considered first, evolves until time  $t_1$  when the second attribute  $k_2$  comes into consideration, triggering a change in the accumulation process. This attribute in turn is considered until time  $t_2$  when a third attribute  $k_3$  is considered and so forth until a decision is initiated (or  $t_L$  is reached). Let the random variables  $T_A$  and  $T_B$  denote the finite time when the process reaches a decision threshold  $\theta_A$  or  $-\theta_B$ , stops, and a decision response for A

or B is initiated. With the switching times  $t_l$  replaced by integers  $n_l \approx t_l/\tau$ , the choice probability  $Pr[\text{choose A}] = Pr(T_A < \infty)$  is then approximated by the value  $p_A$  obtained from the discrete random walk model as

$$Pr(T_A < \infty) \approx p_A := Z' \sum_{i=1}^{n_1} Q_{k_1}^{i-1} R_{A,k_1} + Z' Q_{k_1}^{n_1} \sum_{i=n_1+1}^{n_2} Q_{k_2}^{i-(n_1+1)} R_{A,k_2} + \dots + Z' Q_{k_1}^{n_1} \dots Q_{k_{L-1}}^{n_{L-1}-n_{L-2}} \sum_{i=n_{L-1}+1}^{n_L} Q_{k_L}^{i-(n_{L-1}+1)} R_{A,k_L}, \tag{3}$$

where  $Z$  is the probability distribution for the initial state  $X(0)$ . For instance, for an unbiased process,  $Z$  would be a coordinate vector with probability 1 at state 0 halfway between the decision thresholds. The remaining vectors and matrices are those defined in (2). The evidence accumulation process for a successive attribute starts with the final evidence state of the previous attribute. Note that  $Z' Q_{k_1}^{n_1} \dots Q_{k_{L-1}}^{n_{L-1}-n_{L-2}}$  are defective distributions, i.e., the entries of these vectors do not sum up to 1, for the states of the random walk at discrete times  $n_1, \dots, n_{L-1}$ . Further note that the stochastic process is time homogeneous within each time interval  $[0, t_1)$  to  $[t_{L-1}, t_L]$  but non-homogeneous across  $[0, t_L]$  (see Diederich, 1992, 1995).

Similarly, the mean response time for choosing alternative A is approximated as

$$E[T_A | \text{choose A}] \approx ET_A := \frac{\tau}{p_A} \left[ Z' \sum_{i=1}^{n_1} i Q_{k_1}^{i-1} R_{A,k_1} + Z' Q_{k_1}^{n_1} \sum_{i=n_1+1}^{n_2} i Q_{k_2}^{i-(n_1+1)} R_{A,k_2} + \dots + Z' Q_{k_1}^{n_1} \dots Q_{k_{L-1}}^{n_{L-1}-n_{L-2}} \sum_{i=n_{L-1}+1}^{n_L} i Q_{k_L}^{i-(n_{L-1}+1)} R_{A,k_L} \right]. \tag{4}$$

The probability and the mean response time for choosing alternative B can be determined similarly. Note that  $p_0 := 1 - (p_A + p_B)$ , the probability of not making a decision until the final time  $t_L$ , is strictly positive if  $t_L < \infty$ . As shown in Diederich (1997), these formulas can be further compactified. We will do this below for the general case of deterministic and random schedules by deriving an efficient recursion for their evaluation.

#### 3.2. RANDOM TIME AND ORDER SCHEDULE

The above derivation of formulas for choice probabilities and mean response times for a deterministic time and order schedule have counterparts for random schedules which we describe next in three steps.

##### 3.2.1. Random order schedule

For generating the attribute order  $\{k_l\}_{l=1, \dots, L}$ , we consider stochastic  $K \times K$  matrices  $D^{(l)}$  such that  $d_{k'k}^{(l)} \geq 0$  describes the

probability with which attention switches from the  $k'$ -th attribute to the  $k$ -th attribute at switching time  $t_l \approx \tau n_l$ ,  $l = 1, \dots, L - 1$ . Normally,  $d_{kk}^{(l)} = 0$  would be assumed, to avoid a no switching situation. For two attributes  $K = 2$ , we must then have  $d_{11}^{(l)} = d_{22}^{(l)} = 0$ ,  $d_{12}^{(l)} = d_{21}^{(l)} = 1$ , and the attribute sequence is either  $(1, 2, 1, 2, \dots)$  or  $(2, 1, 2, 1, \dots)$ , depending on whether  $k_1 = 1$  or  $k_1 = 2$ . For three attributes and  $L = 3$ , choosing

$$D^{(1)} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}, \quad D^{(2)} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 3/4 & 1/4 & 0 \end{bmatrix},$$

would for  $k_1 = 1$  result in order sequences  $(1, 2, 1)$ ,  $(1, 3, 1)$ ,  $(1, 3, 2)$  with probability  $1/2$ ,  $3/8$ ,  $1/8$ , respectively. The above matrix  $D^{(1)}$  models the situation when no preference or bias for considering attributes can be asserted.

### 3.2.2. Random time schedule

We assume that the number of discrete time steps during which attention is paid to the  $k$ -th attribute is a discrete random variable denoted by  $T_{at}$  with given distribution. In principle, this distribution may change its type and may have different parameters, such as expected value, depending on the attribute and the attribute order  $\{k_l\}_{l=1, \dots, L}$ . This can be used to model time pressure and other temporal effects. However, often we assume one and the same distribution type for attention times across all attributes, and allow for different parameters only.

For instance, the *geometric distribution* (as implicitly considered in Diederich, 1997) is given by

$$\Pr(T_{at} = n) = (1 - r)^{n-1} r, \quad n = 1, 2, \dots,$$

and characterized by a single parameter  $r > 0$ , with expectation  $1/r$  and variance  $(1 - r)/r^2$ , and the uniform distribution is defined as

$$\Pr(T_{at} = n) = \frac{1}{2M + 1}, \quad n = N - M, \dots, N + M,$$

with parameters  $N$  and  $M = 0, 1, \dots, N - 1$  and expectation  $N$  and variance  $M(M + 1)/3$ . Details for other tested distributions (Poisson with parameter  $\lambda > 0$ , and binomial distributions with parameters  $n$  and  $p$ ) are omitted. For comparable expectation values  $E(T_{at})$  (i.e., for parameter choices  $1/r \approx N \approx \lambda \approx np$ ), the geometric distribution has much larger variance than the Poisson, binomial and uniform distribution with  $M \approx \sqrt{N}$  (the latter are very close to each other). **Figure 3** shows the pdf and cdf for different  $T_{at}$  distributions with fixed mean value  $E(T_{at}) = 300$ . The two uniform distributions are with  $M = 150 = N/2$  and  $M = 299 = N - 1$ . Varying the parameter  $M$  of the uniform distribution allows us to produce intermediate results between the deterministic and geometric distribution cases as shown in the following.

### 3.2.3. Constructing random time and order schedules

We create a *random time and order schedule* of length  $L$  in two steps: First, given an initial distribution of  $k_1 \in \{1, \dots, K\}$ , we

create the attribute sequence  $\{k_l\}_{l=2, \dots, L}$  using a non-stationary Markov chain model with transition probability matrices  $D^{(l)}$ ,  $l = 1, \dots, L - 1$ . In a second step, for each  $l = 1, \dots, L$ , the attention time  $T_{at}^{(l)} = n_l - n_{l-1}$  is created by the discrete random variable responsible for the attention time paid to the  $k_l$ -th attribute, choices are independent for the different  $l$ . Consequently,  $t_l - t_{l-1} \approx \tau T_{at}^{(l)}$  is the real attention time paid to the  $k_l$ -th attribute. We note that *semi-random schedules*, where the sequence  $\{k_l\}$  is given deterministically, and only the  $T_{at}^{(l)}$  are determined as in the second step outlined above, are covered if we choose the  $D^{(l)}$  such that  $d_{k_l, k_{l+1}}^{(l)} = 1$ .

To understand the recursive computation of choice probabilities and mean response times in this more general case, we first consider the special cases  $L = 1, 2$ , and illustrate the derivation on some distribution types of the random variable  $T_{at}$  generating attention times by providing concrete formulas. In general, the distribution for  $T_{at}$  is given by its probability mass distribution (pdf) and cumulative distribution function (cdf)

$$\Pr(T_{at} = n) = p_{n,k}, \tag{5}$$

$$\Pr(T_{at} \leq n) = f_{n,k} := \sum_{i=0}^n p_{i,k}, \quad n = 0, 1, \dots$$

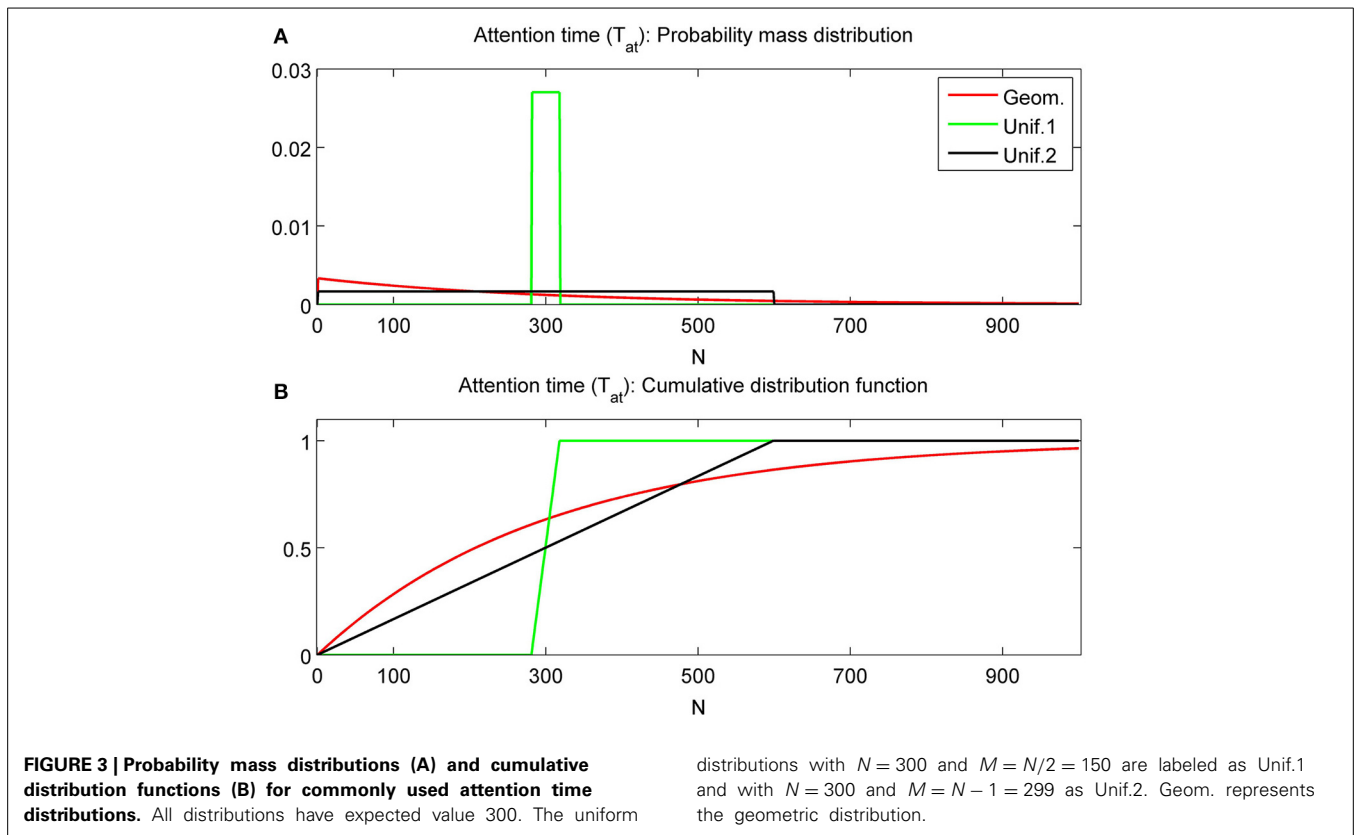
We start with  $L = 1$ , and will drop the index  $l$  from the notation introduced in the previous subsection. Since the probability of choosing alternative  $A$  at the  $i$ -th step is given by  $Z' Q_k^{i-1} R_{A,k}$ ,  $i = 1, \dots, T_{at}$ , and  $T_{at}$  is a random variable distributed according to (5) we get

$$\begin{aligned} p_{A,k} &= \sum_{n=1}^{\infty} p_{n,k} Z' \left( \sum_{i=1}^n Q_k^{i-1} \right) R_{A,k} \\ &= Z' \left[ \sum_{i=1}^{\infty} \left( \sum_{n=i}^{\infty} p_{n,k} \right) Q_k^{i-1} \right] R_{A,k} \\ &= Z' \left[ \sum_{i=0}^{\infty} (1 - f_{i,k}) Q_k^i \right] R_{A,k}. \end{aligned}$$

A similar formula holds for  $p_{B,k}$ . To avoid repetition, introduce the row vector  $p_{AB,k} := [p_{B,k}, p_{A,k}]$ , then

$$p_{AB,k} = Z' V_k, \quad V_k := \left[ \sum_{i=0}^{\infty} (1 - f_{i,k}) Q_k^i \right] R_k. \tag{6}$$

The  $2 \times (m - 2)$  matrix  $V_k$  depends on the attribute and its parameters via  $Q_k, R_k$ , and on the chosen attention time distribution and the cdf  $(f_{n,k})$ . For the discussed concrete attention time distributions these matrices may be precomputed, in some cases closed-form expressions can be found, e.g., for the geometric distribution with parameter  $r = r_k$  we have



$$V_k = \sum_{i=0}^{\infty} \left( \sum_{j=i+1}^{\infty} r_k (1 - r_k)^{j-1} \right) Q_k^i R_k$$

$$= \sum_{i=0}^{\infty} (1 - r_k)^i Q_k^i R_k = (I - (1 - r_k)Q_k)^{-1} R_k.$$

Next we discuss choice probabilities for the case  $L = 2$ , assuming for simplicity that the attention time distribution is the same for all attributes. To save on indices, denote  $k_1 \equiv k'$ ,  $k_2 \equiv k$ , and  $D^{(1)} \equiv D$  (this matrix is responsible for the random choice of  $k$  given any  $k'$ ). Then the decision probability vector  $p_{AB,k',k}$  for reaching alternatives  $B$  or  $A$  in with attribute order  $(k', k)$  has two parts: the probabilities of having decided on while still considering the  $k'$ -th attribute (i.e.,  $T_A/\tau \leq T'_{at}$ , where  $T'_{at}$  is the randomly generated attention time for the first attribute  $k'$ ) plus the probabilities that  $\tau T'_{at} < T_A/\tau \leq T'_{at} + T_{at}$ , where  $T_{at}$  is the randomly (and independently) generated attention time for the second attribute  $k$ . On top of this,  $k$  itself is randomly chosen according to the entries in the  $k'$ -th row of  $D$ . Thus, for each fixed  $k_1 = k'$  and  $n_1 = T'_{at}$  according to (6) probabilities for reaching a decision after  $n_1$  are given by

$$\left[ \Pr \left( T'_{at} < \frac{T_B}{\tau} < \infty \right), \Pr \left( T'_{at} < \frac{T_A}{\tau} < \infty \right) \right]_{n_1 = T'_{at}, k_1 = k'}$$

$$\approx \sum_{k=1}^K d_{k'k} Z' Q_k^{n_1} V_k = Z' Q_{k'}^{T'_{at}} \left( \sum_{k=1}^K d_{k'k} V_k \right).$$

Thus, for  $L = 2$ , the choice probabilities (under the assumption that  $k_1 = k'$  is fixed) can be obtained as

$$[p_B, p_A]_{k_1 = k'} = Z' V_{k'} + \sum_{n \geq 0} p_{n,k'} Z' Q_{k'}^n \left( \sum_{k=1}^K d_{k'k} V_k \right)$$

$$= Z' \left[ V_{k'} + \left( \sum_{n \geq 0} p_{n,k'} Q_{k'}^n \right) \left( \sum_{m=1}^M d_{k'm} V_m \right) \right]$$

$$= Z' \left[ V_{k'} + B_{k'} \left( \sum_{k=1}^K d_{k'k} V_k \right) \right], \quad k' = 1, \dots, K,$$

where

$$B_k = \sum_{n \geq 0} p_{n,k} Q_k^n, \quad k = 1, \dots, K, \tag{7}$$

are  $(m - 2) \times (m - 2)$  matrices depending on the attribute and attention time distribution type. For example, for the geometric distribution this simplifies to  $B_k = r_k Q_k (I - (1 - r_k) Q_k)^{-1}$ , closed form expressions are available for Poisson, binomial, and uniform distributions as well.

For arbitrary  $L$ , it is more convenient to write the resulting recursion in terms of block-matrix-vector operations. Denote by

- Z** the  $K \times 1$  array with each entry equal to the initial distribution  $Z$  (and think of  $\mathbf{Z}'$  as its transpose, a  $1 \times K$  array with entries  $Z'$ ),
- B** the  $K \times K$  diagonal array with the  $B_k$  on the diagonal (similarly for **C** defined later),
- I** the  $K \times K$  diagonal array, with identity matrices  $I$  of the appropriate size on the diagonal,
- V** the  $K \times 1$  array with the  $V_k$  as entries (similarly for **W** defined later), and
- p<sub>AB</sub>** the  $K \times 2$  matrix, whose rows are the choice probabilities  $[p_A, p_B]_{|k_1=k}$  defined before in the case  $L = 2$ .

Then the above result for  $L = 2$  can be compactly written as

$$\mathbf{p}_{AB} = \mathbf{Z}' (\mathbf{I} + \mathbf{B}D) \mathbf{V}. \tag{8}$$

Note that the product  $\mathbf{B}D$  of the array **B** with the matrix  $D$  is interpreted as the  $K \times K$  array with  $d_{k'k}B_{k'}$  as entry in row  $k'$  and column  $k$ . Moreover, by iterating (8), one arrives at the formula for arbitrary  $L$ :

$$\mathbf{p}_{AB} = \mathbf{Z}' (\mathbf{I} + \mathbf{B}D^{(1)}) \dots (\mathbf{I} + \mathbf{B}D^{(L-1)}) \mathbf{V}. \tag{9}$$

Formulas for mean response times can be derived similarly. Indeed, for  $L = 1$ , denote by  $ET_{A,k}$  the mean response time for reaching alternative  $A$  when considering the  $k$ -th attribute for a random time  $T_{at}$  distributed according to (5). Then  $ET_{A,k} \approx \tau et_{A,k}/p_{A,k}$ , where

$$\begin{aligned} et_{A,k} &= \sum_{n=1}^{\infty} p_{n,k} \left( \sum_{i=0}^{n-1} (i+1)Z'Q_k^i \right) R_{A,k} \\ &= Z' \left[ \sum_{i=0}^{\infty} \left( \sum_{n=i+1}^{\infty} p_{n,k} \right) (i+1)Q_k^i \right] R_{A,k} \\ &= Z' \left[ \sum_{i=0}^{\infty} (1 - f_{i,k})(i+1)Q_k^i \right] R_{A,k}. \end{aligned} \tag{10}$$

Similarly for  $ET_{B,k}$  and  $et_{B,k}$ . Thus, similar to (6), we can write

$$et_{AB,k} := [et_{B,k}, et_{A,k}] = Z'W_k, \tag{11}$$

$$W_k := \left[ \sum_{i=0}^{\infty} (1 - f_{i,k})(i+1)Q_k^i \right] R_k, \quad k = 1, \dots, K.$$

The matrices  $W_k$  can be precomputed to any accuracy at essentially the same cost as the  $V_k$ . For particular distributions, the formulas can be turned into closed form expressions.

Next, let us look at  $L = 2$ . By using similar notation and arguments as for choice probabilities, the quantities  $et_{A,k',k}$ ,  $et_{B,k',k}$  have a part before and after  $T'_{at}$ . This, together with (10), (11), gives

$$\begin{aligned} et_{AB|k_1=k'} &= Z'W_{k'} + \sum_{n=0}^{\infty} p_{n,k}Z'Q_{k'}^n \left( \sum_{k=1}^K d_{k'k}(nV_k + W_k) \right) \\ &= Z' \left[ W_{k'} + \left( \sum_{i=0}^{\infty} p_{i,k'}iQ_{k'}^i \right) \left( \sum_{k=1}^K d_{k'k}V_k \right) \right. \\ &\quad \left. + \left( \sum_{i=0}^{\infty} p_{i,k'}Q_{k'}^i \right) \left( \sum_{k=1}^K d_{k'k}W_k \right) \right] \\ &= Z' \left[ W_{k'} + C_{k'} \left( \sum_{k=1}^K d_{k'k}V_k \right) + B_{k'} \left( \sum_{k=1}^K d_{k'k}W_k \right) \right], \end{aligned}$$

where

$$C_k = \sum_{n \geq 0} p_{n,k}nQ_k^n, \quad k = 1, \dots, K. \tag{12}$$

Thus, the counterpart of (8) is

$$\mathbf{et}_{AB} = \mathbf{Z}'((\mathbf{C}D)\mathbf{V} + (\mathbf{I} + \mathbf{B}D)\mathbf{W}), \tag{13}$$

From here, combining with (8), a joint recursion for computing  $\mathbf{p}_{AB}$  and  $\mathbf{et}_{AB}$  results:

$$\begin{aligned} [\mathbf{p}_{AB}, \mathbf{et}_{AB}] &= [\mathbf{Z}', \mathbf{Z}'] \begin{bmatrix} \mathbf{I} + \mathbf{B}D^{(1)} & \mathbf{0} \\ \mathbf{C}D^{(1)} & \mathbf{I} + \mathbf{B}D^{(1)} \end{bmatrix} \dots \\ &\quad \begin{bmatrix} \mathbf{I} + \mathbf{B}D^{(L-1)} & \mathbf{0} \\ \mathbf{C}D^{(L-1)} & \mathbf{I} + \mathbf{B}D^{(L-1)} \end{bmatrix} \begin{bmatrix} \mathbf{V} \\ \mathbf{W} \end{bmatrix}. \end{aligned} \tag{14}$$

We conclude this section with a few remarks. In Diederich (1997), under the name MADD/pp, a slightly different presentation of random schedules is given for the special case of geometrically distributed attention times. It is not hard to see, that (with the notation  $r_{ij}$  used in the  $K = 3$  example presented in Section 4.2 Diederich, 1997) our model is equivalent to MADD/pp as  $L \rightarrow \infty$ , if we set  $r_k = 1 - r_{kk}$  for the parameters  $r$  of the geometrically distributed  $T_{at}$ ,  $k = 1, 2, 3$ , and  $d_{kk} = 0$ ,  $d_{kk'} = r_{kk'}/(1 - r_{kk})$ ,  $k' \neq k$ , for the entries of the matrix  $D = D^{(l)}$ ,  $l \geq 1$ . The advantage of the MADD/pp model is that it provides closed form formulas for the case  $L = \infty$ , a possibility that we did not pursue here for other types of attention time distributions.

In previous sequential decision models with finite  $L$  (Diederich, 1997), the last attribute was always considered infinitely long (infinite decision horizon) to avoid the situation of no decision, i. e.,  $p_0 > 0$ . This can be incorporated into the current model by modifying the definition of the matrices  $V_k$ ,  $W_k$  corresponding to the last interval  $[t_{L-1}, \infty)$  to

$$V_k = (I - Q_k)^{-1}R_k, \quad W_k = (I - Q_k)^{-2}R_k, \quad k = 1, \dots, K,$$

and modifying the recursion (14) slightly. Alternatively, one can artificially change the parameters of the attention time distribution for  $l = L$  such that its expected value is sufficiently large, and make  $p_0$  practically negligible. Since infinite decision horizons do not seem to adequately reflect the situation of a real decision process or laboratory experiment, it might be interesting to work under scenarios where  $t_L$  is fixed and finite that we described in this paper.

## 4. SIMULATIONS

We present some simulations that demonstrate the predictive power of the proposed model. We focus on features that have not been considered in Diederich (1997) for the deterministic case. Throughout this section we fix certain parameters, such as  $\sigma = 1$ ,  $\theta_A = -\theta_B = 10$ ,  $\Delta = \frac{1}{4}$ ,  $\tau = \frac{1}{16}$  (this implies a state space size of  $m = 81$ ), and always start at the neutral position  $X(0) = 0$  between choice alternatives  $A$  and  $B$ .

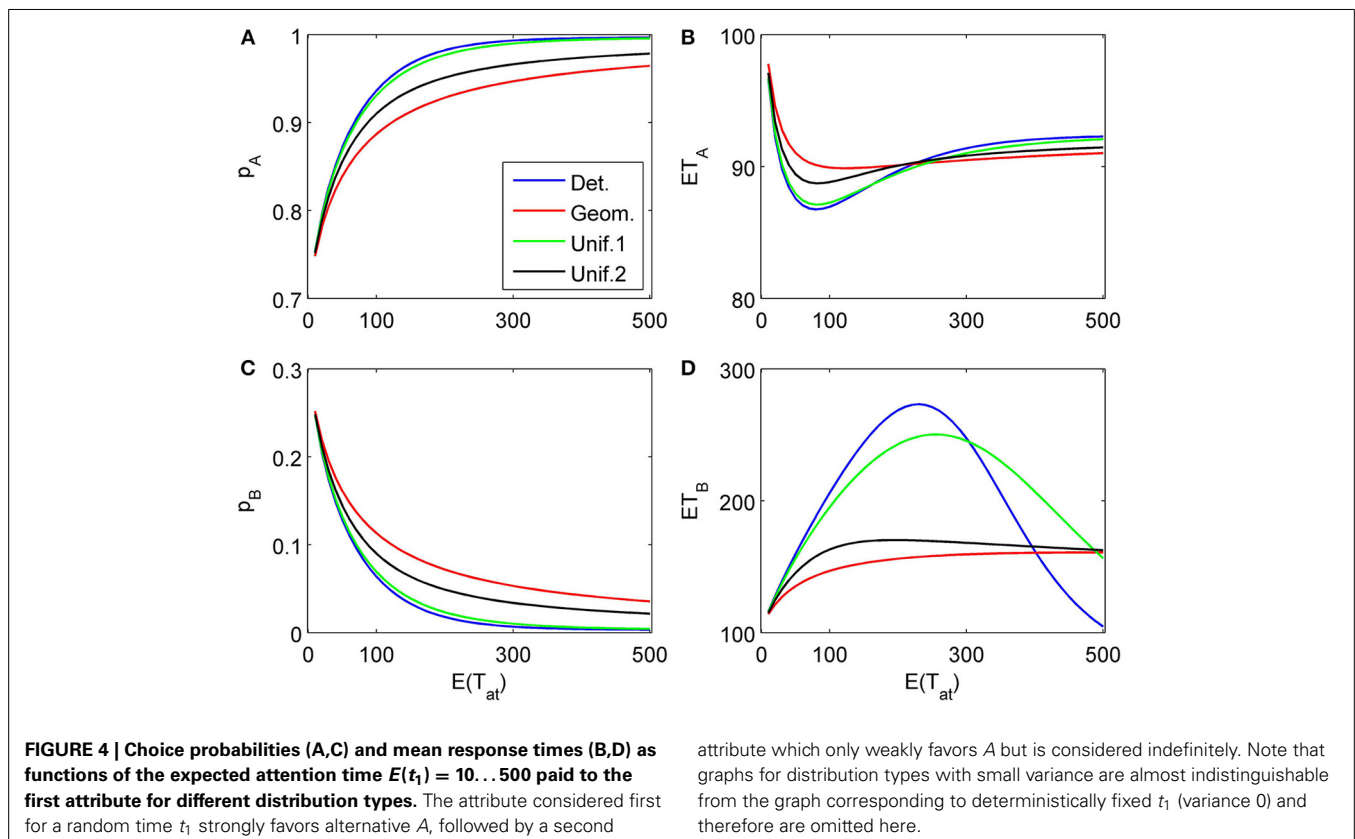
### 4.1. IMPACT OF ATTENTION TIME DISTRIBUTIONS

First, we show how different assumptions on the randomness of the attention time  $T_{at}$  (i.e., the time spent on considering a certain attribute) influences choice probabilities and mean response times. In the first example, we assume just two attributes with parameters  $\delta_1 = 0.2$ ,  $\gamma_1 = 0.03$ ,  $\delta_2 = 0.04$ ,  $\gamma_2 = 0.003$ , both attributes favor alternative  $A$ , the first one more strongly than the second one<sup>2</sup>. The attributes are considered only once ( $L = 2$ ), with order  $k_1 = 1$ ,  $k_2 = 2$ . The first attribute is considered for time  $t_1 = \tau n_1$ , where  $n_1$  is a random variable  $T_{at}$  described above with given expectation  $N$ . For the second

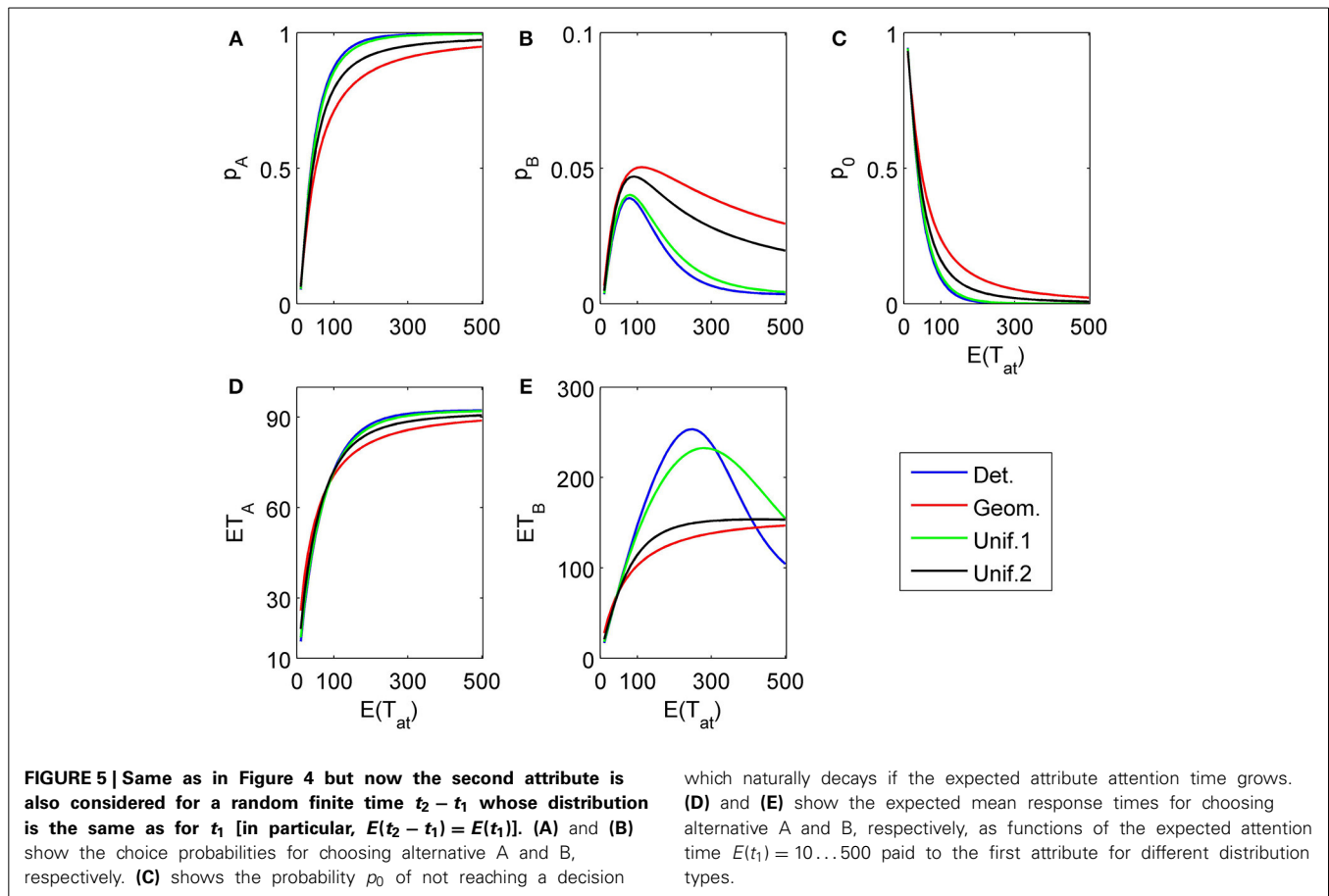
<sup>2</sup>Note that when looking only at the numerical values of the drift parameter  $\delta_1 = 0.2$  and the decision criterion  $\theta_A = 10$  and assuming that the attention times  $t_1$  to the first attribute are large enough it would suggest mean response times in the range  $T_A \approx 50$  (and very small  $p_B$ ). However, since  $\gamma_1 = 0.03$  it leads to a negative effective drift  $\delta_1 - \gamma_1 X(t)$  if  $X(t)$  comes close  $\theta_A$ , and the mean response times become much longer. This also demonstrates the effect of the parameter  $\gamma_k$ , and a difference between Ornstein-Uhlenbeck process and Wiener process based models.

attribute we compare two situations: (1) We assume an infinitely long decision horizon  $t_2 = \infty$ , and (2) we determine a finite time horizon  $t_2 = \tau n_2$  by choosing  $n_2 = n_1 + T_{at}$  which is also  $T_{at}$  distributed with the same expected value  $N$ . These two situations are depicted in Figures 4, 5. The graphs show choice probabilities and mean response times as functions of the expectation  $\tau E(T_{at})$  of the real attention times. Lines of different color represent different distributions. Distributions with a small variance, such as the Poisson distribution, the binomial distribution, and the uniform distribution with  $M \approx \sqrt{N}$  produce results indistinguishable from the deterministic case. This holds for all tested situations shown below. This means, small uncertainties in attention time spans do not influence the observable choice frequencies and mean response times. However, as the variance of the attention times grows, we see quantitative and qualitative changes. Compared to the deterministic attention time situation, the geometric distribution differs most, and the uniform distributions with  $M = N/2 = 150$  (Unif.1) and  $M = N - 1 = 299$  (Unif.2) are intermediate. Moreover, there is expectedly a big difference for small mean attention times between finite and infinite decision horizons. Most importantly, for the former case it predicts a probability  $p_0 > 0$  of not deciding within the available time  $t_2$ . We claim that for many situations, where an infinite time horizon does not represent reality well enough, our finite schedule model might be more appealing. This aspect will be pursued in further research.

Figures 6, 7 show similar simulation results for the situation of considering first an attribute favoring  $B$  ( $\delta_1 = -0.1$ ,  $\gamma_1 = 0$ )







followed by an attribute more strongly favoring A ( $\delta_2 = 0.2$ ,  $\gamma_2 = 0.03$ ). As expected, the results look now different, however, the main conclusions from the previous example concerning the influence of the randomness type for attention times and the differences for finite vs. infinite time horizons remain the same. Most importantly here, the model predicts a preference reversal (i.e., choice probabilities from below 0.5 to above 0.5) as a function of attention time when one attribute is in favor of choosing alternative A and the other in favor of choosing alternative B. Parameter studies, as in Diederich (1997), will be pursued further elsewhere.

To complete the picture, we show a three-attribute example ( $K = 3$ ) in Figure 8. The chosen attribute parameters are now  $\delta_1 = 0.04$ ,  $\gamma_1 = 0.003$ ,  $\delta_2 = -0.1$ ,  $\gamma_2 = 0$ ,  $\delta_3 = 0.2$ ,  $\gamma_3 = 0.03$ , i.e., a weakly in favor of A, in favor of B, and strongly in favor of A sequence of attributes. Attention times for the first two attributes are chosen independently from each other but with the same distribution with fixed mean value; the last attribute is considered indefinitely.

#### 4.2. DEPENDENCE ON ATTRIBUTE ORDER

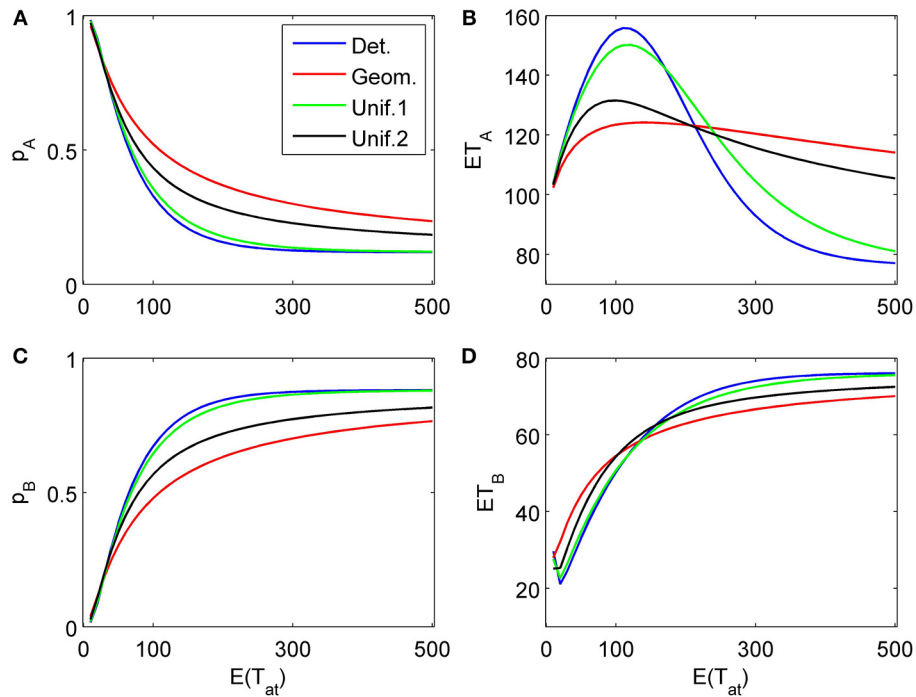
The proposed sequential decision model is sensitive to the order in which the attributes are considered. If we consider in the aforementioned second two-attribute example the attribute in favor of A first, and then the attribute in favor of B we get very different patterns as shown in Figure 9 compared to Figure 6. A similar effect is true for the above  $K = 3$  example. In Figure 10, the

attribute in favor of B is now the last one; the graphs need to be compared with Figure 8. One interesting pattern can be observed. If the evidence for choosing one alternative decreases in the sequence of attribute consideration then the model predicts faster choice response times for the more frequently chosen alternative—a typical pattern observed in response time analysis. However, if the evidence increases in the sequence of attribute consideration then the model predicts faster choice response times for the less frequently chosen alternative which has been called *fast error*, as shown in Figure 11 compared to Figure 4. Simply by changing the order of attribute processing the model predicts a complex pattern of choice response times and choice probabilities.

So far, all examples shown are with a fixed, deterministic attribute order with no repetitions (semi-random schedule,  $L = K$ ). The evaluation of fully random time and order schedules requires larger  $L$ , and will be presented elsewhere.

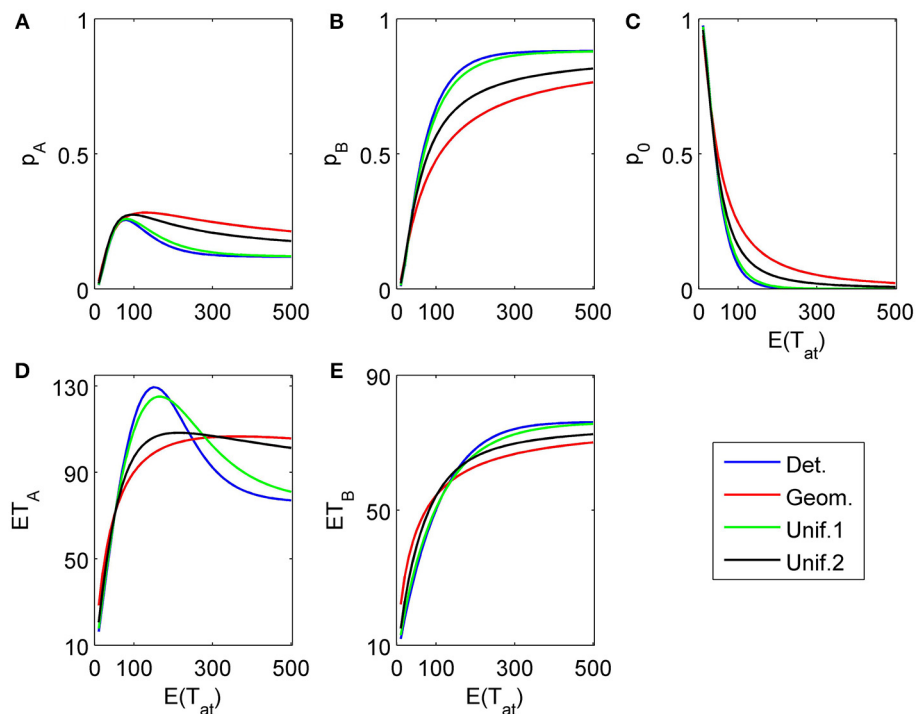
#### 5. CONCLUDING REMARKS

The proposed *multiattribute attention switching* (MAAS) model can predict a very complex choice probability/(mean) choice response time pattern. It may appear too flexible to be testable. However, this is not the case. If two attributes both favor alternative, A say, and the first attribute that is considered provides more evidence for choosing A than the second ( $\delta_1 > \delta_2$ ), then the model predicts always shorter response times for the more



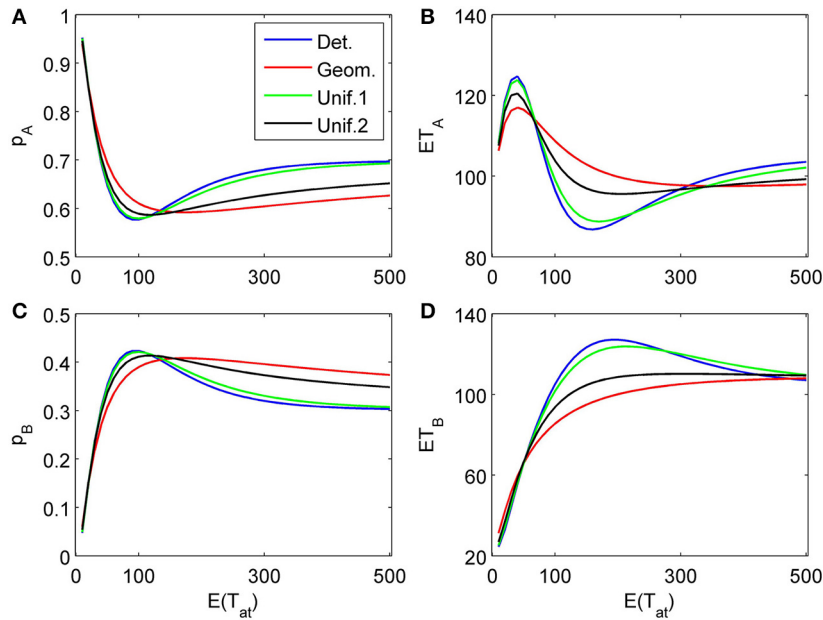
**FIGURE 6 | Choice probabilities (A,C) and mean response times (B,D) for a decision situation where an attribute favoring alternative B is considered first for a random time  $t_1$ , followed by a second attribute strongly favoring A but considered indefinitely.** We show graphs of

choice probabilities and mean response times as functions of the expected attention time  $E(t_1) = 10 \dots 500$  paid to the first attribute for different distribution types. Again, graphs for distribution types with small variance are indistinguishable from each other.



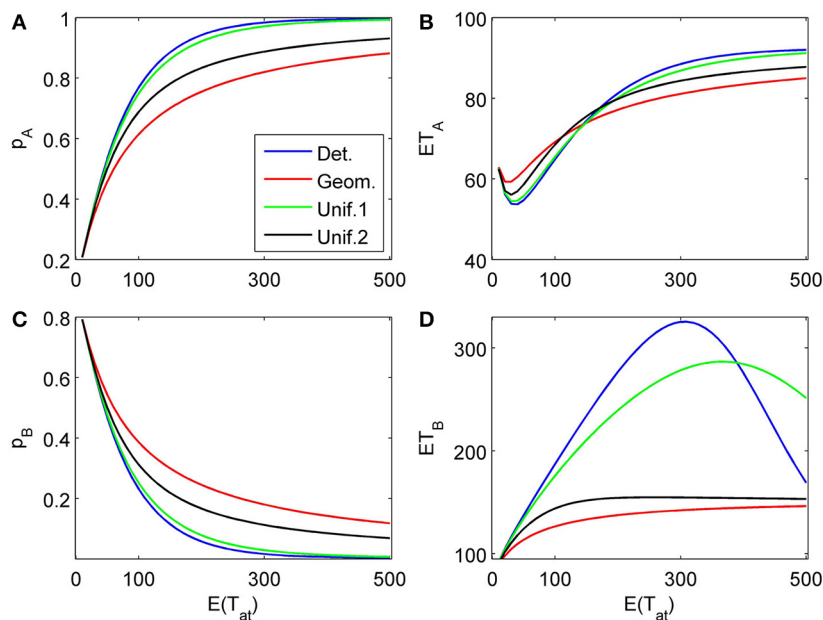
**FIGURE 7 | Same as in Figure 6 but now the second attribute is also considered for a random finite time  $t_2 - t_1$  whose distribution is the same as for  $t_1$ .** (A), (B), and (C) show the choice probabilities

for choosing alternatives A, B and none, respectively. (D) and (E) show the mean response times for choosing alternatives A and B, respectively.



**FIGURE 8 | Choice probabilities (A,C) and mean response times (B,D) for a decision model with three attributes.** An attribute weakly favoring alternative A is considered first for a random time  $t_1$ , followed by a second attribute favoring B considered for a random time  $t_2 - t_1$ , while the last attribute (strongly favoring A) is considered indefinitely. The random attention

times  $t_1$  and  $t_2 - t_1$  for the first two attributes are independently chosen from the same distribution. We show graphs of choice probabilities and mean response times as functions of the expected attention time  $E(t_1) = E(t_2 - t_1) = 10 \dots 500$  for different distribution types. Again, small variance distributions yield almost identical results.

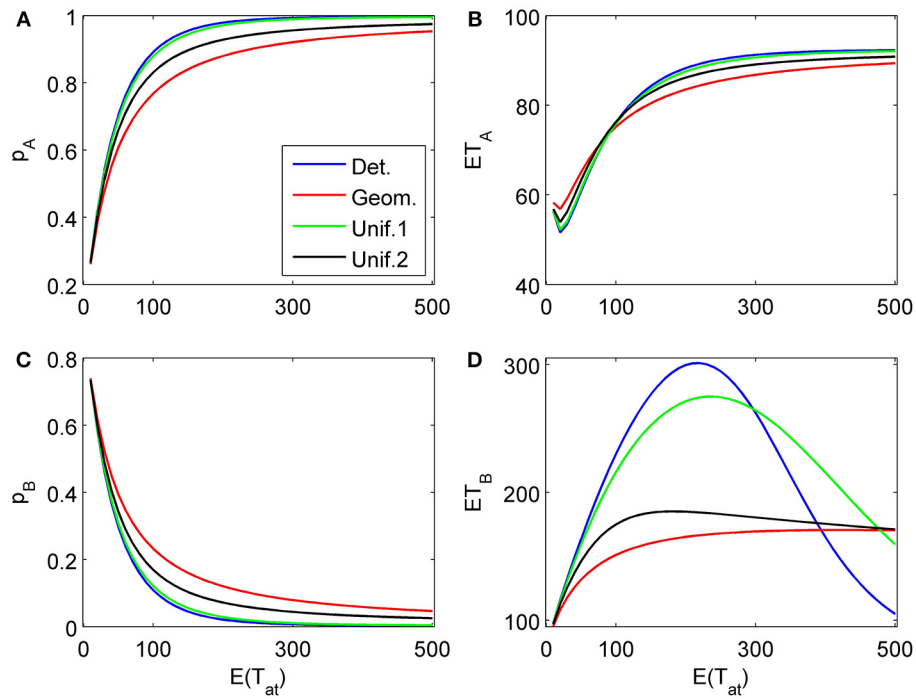


**FIGURE 9 | Same as in Figure 6 but with a different attribute order: First the attribute strongly in favor of A is considered for a finite random time  $t_1$ , then the attribute favoring B is considered indefinitely long.** (A) and

(C) show the choice probabilities for choosing alternatives A and B respectively. (B) and (D) show the mean response times for choosing alternatives A and B, respectively.

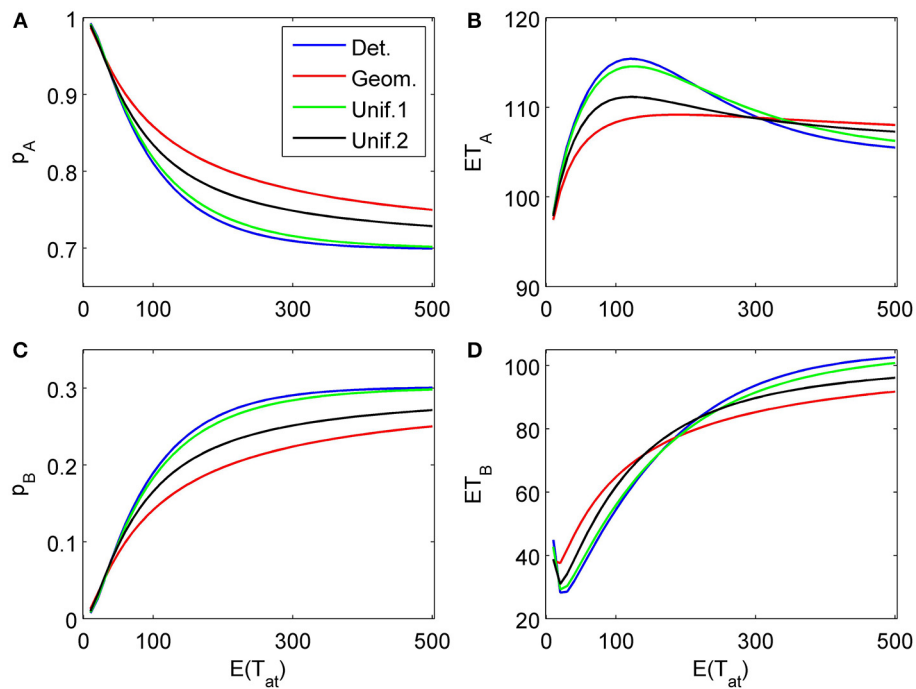
frequently chosen alternative, here A, regardless of the assumed underlying attention time distribution. If the order of processing these attributes is reversed, i.e., the attribute that favors alternative A less is considered first ( $\delta_2 > \delta_1$ ), then the model always

predicts faster responses for the less frequently chosen alternative, here B, again regardless of the assumed underlying attention time distribution. A single stage process can only account for this pattern by assuming variability in starting positions and



**FIGURE 10 |** Same as in Figure 8 but with a different attribute order: First the two attributes in favor of A (strong followed by weak) are considered for finite random periods of time, then the attribute favoring

**B is considered indefinitely long. (A) and (C)** show the choice probabilities for choosing alternatives A and B, respectively. **(B) and (D)** show the mean response times for choosing alternatives A and B, respectively.



**FIGURE 11 |** Same as in Figure 4 but with a different attribute order: The attribute considered first for a random time  $t_1$  weakly favors alternative A, followed by a second attribute which strongly favors A but is

**considered indefinitely. (A) and (C)** show the choice probabilities for choosing alternatives A and B respectively. **(B) and (D)** show the mean response times for choosing alternatives A and B, respectively.

variability in drift rates, i.e., a statistical means where the drift rate itself is a random variable. It is difficult experimentally to disentangle the variability stemming from the stochastic process itself and the variability from the distribution of different drift rates. As Jones and Dzhafarov (2013) pointed out, the predictions of various sequential sampling models rest upon the assumptions made about the assumed probability distributions. This is not the case here. The model is falsifiable without assuming specific distributions. Rather than relying on statistical mechanisms to ensure an observed response patterns we rely on assumptions about cognitive processes such as attention switching and salience. The specific attention time distribution used for an application may be related to the experimental paradigm. For instance, when tracking eye movements, the sequence of attribute consideration and the switching times are directly observable, and a deterministic or a uniform distribution with a small variance is advisable. When all attributes are shown simultaneously, like in complex objects, and attention may shift at any moment in time a geometric distribution or a uniform distribution with a large variance may describe the situation better. Testing the model rigorously will be pursued in the future.

## REFERENCES

- Ashby, F. (1983). A biased random walk model for two choice reaction times. *J. Math. Psychol.* 27, 277–297. doi: 10.1016/0022-2496(83)90011-1
- Ashby, F. (2000). A stochastic version of general recognition theory. *J. Math. Psychol.* 44, 310–329. doi: 10.1006/jmps.1998.1249
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* 113, 700–765. doi: 10.1037/0033-295X.113.4.700
- Busemeyer, J. and Goldstein, W. (1992). Linking together different measures of preference: a dynamic model of matching derived from decision field theory. *Organ. Behav. Hum. Decis. Process.* 52, 370–396. doi: 10.1016/0749-5978(92)90026-4
- Busemeyer, J., and Townsend, J. (1993). Decision field theory: a dynamic-cognitive approach to decision-making in an uncertain environment. *Psychol. Rev.* 100, 432–459. doi: 10.1037/0033-295X.100.3.432
- Churchland, A., Kiani, R., and Shadlen, M. (2008). Survey of decision field theory. *Nat. Neurosci.* 11, 693–702. doi: 10.1038/nn.2123
- Diederich, A. (1992). *Intersensory Facilitation: Race, Superposition, and Diffusion Models for Reaction Time to Multiple Stimuli*. Frankfurt am Main: Verlag Peter Lang.
- Diederich, A. (1995). Intersensory facilitation of reaction time: evaluation of counter and diffusion coactivation models. *J. Math. Psychol.* 39, 197–215. doi: 10.1006/jmps.1995.1020
- Diederich, A. (1997). Dynamic stochastic models for decision making with time constraints. *J. Math. Psychol.* 41, 260–274. doi: 10.1006/jmps.1997.1167
- Diederich, A. (2003). Decision making under conflict: decision time as a measure of conflict strength. *Psychon. Bull. Rev.* 10, 167–176. doi: 10.3758/BF03196481
- Diederich, A. (2008). A further test on sequential sampling models accounting for payoff effects on response bias in perceptual decision tasks. *Percept. Psychophys.* 70, 229–256. doi: 10.3758/PP.70.2.229
- Diederich, A., and Busemeyer, J. (1999). Conflict and the stochastic dominance principle of decision making. *Psychol. Sci.* 10, 353–359. doi: 10.1111/1467-9280.00167
- Diederich, A., and Busemeyer, J. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time and simple response time. *J. Math. Psychol.* 47, 304–322. doi: 10.1016/S0022-2496(03)00003-8
- Diederich, A., and Busemeyer, J. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: threshold-bound, drift rate-change, or two-stage-processing hypothesis. *Percept. Psychophys.* 68, 194–207. doi: 10.3758/BF03193669
- Ditterich, J. (2006). Stochastic models of decisions about motion direction: behavior and physiology. *Neural Netw.* 19, 981–1012. doi: 10.1016/j.neunet.2006.05.042
- Edwards, W. (1965). Optimal strategies for seeking information: models for statistics, choice reaction times, and human information processing. *J. Math. Psychol.* 2, 312–329. doi: 10.1016/0022-2496(65)90007-6
- Gao, J., Tortell, R., and McClelland, J. L. (2011). Dynamic integration of reward and stimulus information in perceptual decision-making. *PLoS ONE* 6:e16749. doi: 10.1371/journal.pone.0016749
- Gold, J., and Shadlen, M. (2007). The neural basis of decision making. *Ann. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038
- Heath, R. (1981). A tandem random walk model for psychological discrimination. *Br. J. Math. Stat. Psychol.* 34, 76–92. doi: 10.1111/j.2044-8317.1981.tb00619.x
- Johnson, J., and Busemeyer, J. (2005). A dynamic, stochastic, computational model of preference reversal phenomena. *Psychol. Rev.* 112, 841–861. doi: 10.1037/0033-295X.112.4.841
- Jones, M., and Dzhafarov, E. N. (2013). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychol. Rev.* 121, 1–32. doi: 10.1037/a0034190
- Laming, D. (1968). *Information Theory of Choice Reaction Times*. Oxford: Academic Press.
- Link, S., and Heath, R. (1975). A sequential theory of psychological discrimination. *Psychometrika* 40, 77–105. doi: 10.1007/BF02291481
- Nosofsky, R., and Palmeri, T. (1997). An exemplar based random walk model of speeded classification. *Psychol. Rev.* 104, 266–300. doi: 10.1037/0033-295X.104.2.266
- Pike, A. (1973). Response latency models for signal detection. *Psychol. Rev.* 80, 53–68. doi: 10.1037/h0033871
- Pleskac, T., and Busemeyer, J. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Acta Neurobiol. Exp.* 117, 864–901. doi: 10.1037/a0019737
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108. doi: 10.1037/0033-295X.85.2.59
- Ratcliff, R., and Smith, P. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychol. Rev.* 111, 333–367. doi: 10.1037/0033-295X.111.2.333
- Rorie, A., Gao, J., McClelland, J., and Newsome, W. (2010). Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (lip) of the macaque monkey. *PLoS ONE* 5:e9308. doi: 10.1371/journal.pone.0009308
- Stone, M. (1960). Models for choice-reaction time. *Psychometrika* 25, 251–260. doi: 10.1007/BF02289729
- Usher, M., and McClelland, J. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* 108, 550–592. doi: 10.1037/0033-295X.108.3.550
- Van Zandt, T., Colonius, H., and Proctor, R. (2000). A comparison of two reaction time models applied to perceptual matching. *Psychon. Bull. Rev.* 7, 208–256. doi: 10.3758/BF03212980
- Voss, A., Nagler, M., and Lerche, V. (2013). Diffusion models in experimental psychology. *Exp. Psychol.* 60, 385–402. doi: 10.1027/1618-3169/a000218

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 07 April 2014; accepted: 19 August 2014; published online: 09 September 2014.

Citation: Diederich A and Oswald P (2014) Sequential sampling model for multiattribute choice alternatives with random attention time and processing order. *Front. Hum. Neurosci.* 8:697. doi: 10.3389/fnhum.2014.00697

This article was submitted to the journal *Frontiers in Human Neuroscience*.

Copyright © 2014 Diederich and Oswald. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.