The 2*N*-ary choice tree model for *N*-alternative preferential choice

Lena M. Wollschläger and Adele Diederich*

School of Humanities and Social Sciences, Jacobs University Bremen, Bremen, Germany

*Correspondence:

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

The 2*N*-ary choice tree model accounts for response times and choice probabilities in multi-alternative preferential choice. It implements pairwise comparison of alternatives on weighted attributes into an information sampling process which, in turn, results in a preference process. The model provides expected choice probabilities and response time distributions in closed form for optional and fixed stopping times. The theoretical background of the 2*N*-ary choice tree model is explained in detail with focus on the transition probabilities that take into account constituents of human preferences such as expectations, emotions, or socially influenced attention. Then it is shown how the model accounts for several context-effects observed in human preferential choice like similarity, attraction, and compromise effects and how long it takes, on average, for the decision. The model is extended to deal with more than three choice alternatives. A short discussion on how the 2*N*-ary choice tree model differs from the multi-alternative decision field theory and the leaky competing accumulator model is provided.

Keywords: 2*N*-ary choice tree model, preferential choice, multiple choice alternatives, multi-attribute choice alternatives, response times, choice probabilities, elimination of choice alternatives, computational model

1. INTRODUCTION

Life is full of decisions: Be it the selection of clothing in the morning or of menu for lunch, the question which car to buy or if taking cold medication is necessary. This type of decisions is called preferential choice and has been subject of numerous investigations within the field of decision theory (Koehler and Harvey, 2007, for a review). Several effects have been observed when the decision maker has more than two-choice options (multi-alternative preferential choice). Hick's Law (Hick, 1952; Hyman, 1953), originally defined in the context of stimulus detection paradigms, postulates a dependency of deliberation time on the number of alternatives. In particular, it states that a linear increase of the number of equally attractive alternatives to choose from leads to a logarithmic increase of the time that passes until the decision is made. Furthermore, a decision maker who is indifferent between two-choice alternatives from a given choice set may change the preference for one or the other alternative when the choice set is enlarged, i.e., the local context may affect the decision and generate preference reversals. Similarity effects (Tversky, 1972), attraction effects (Huber et al., 1982), and compromise effects (Simonson, 1989), for instance, depend on a third alternative that is added to a choice set of two equally attractive but dissimilar alternatives. If the third alternative is very similar to one of the others, the two similar alternatives share their choice frequency and are both chosen less often than the dissimilar one (similarity effect). If the third alternative is similar to one of the others but slightly inferior, it promotes the similar one and increases its choice frequency compared to the dissimilar one (attraction effect). If the third alternative is a compromise between the other two, the decision maker will prefer the compromise to the other alternatives (compromise effect). Besides those preference reversals that emerge from local context,

there might also be influence from background context (Tversky and Simonson, 1993) like a reference point outside of the choice set which – together with the loss-aversion principle (Kahneman and Tversky, 1979) – affects evaluation of the given alternatives.

One challenge for (cognitive) modelers is to think of a model which predicts decision making behavior for multi-alternative preferential choice tasks in general but also accounts for all the aforementioned effects. Another challenge is to formulate the model such that (expected) response times and choice probabilities can be calculated and the model parameters conveniently estimated from the observed choice times and choice frequencies.

Decision field theory (DFT, Busemeyer and Townsend, 1992, 1993) and its multi-attribute extension (Diederich, 1997) predict choice response times and choice probabilities for binary choice tasks. Both approaches provide closed form solutions to calculate these entities. Since then, several attempts have been made to extend this kind of models to multi-alternative preferential choice tasks: multi-alternative DFT (Roe et al., 2001) and the leaky competing accumulator (LCA) model (Usher and McClelland, 2001, 2004) predict choice probabilities for three alternative choice tasks but cannot account for optional choice times, i.e., the time the decision maker needs to come to a decision. Both approaches, however, do account for fixed stopping times, i.e., for an externally determined time limit. Furthermore, multi-alternative DFT and the LCA model both account for the similarity, attraction, and compromise effects using computer simulations to predict the patterns. To do so Roe et al. (2001) interpret DFT as a connectionist network and implement distance-dependent inhibition between the alternatives. Usher and McClelland (2001, 2004) add insights from perceptual choice and neuropsychology to the multi-alternative DFT and propose for their LCA model direct implementation of

loss-aversion by means of an asymmetric value function and global inhibition instead of distance-dependent inhibition.

Our $2N$-ary choice tree model builds on the previous approaches and tries to overcome some of their problems. It is a general model for choice probabilities and response times in choice between N alternatives with D attributes. As such, it provides a way to calculate expected response times, response time distributions, and choice probabilities in closed form by determining the time course of an information sampling process via a random walk on a specific tree. It is able to account for similarity, attraction, and compromise effects which have been most challenging for previous models. In contrast to previous approaches, the $2N$-ary choice tree model accounts for these effects without additional mechanisms like inhibition or loss-aversion and is thus more parsimonious. However, it is possible to implement these mechanisms if the situation requires it.

First, we describe the structure of the $2N$-ary choice tree and the implementation of the random walk on it in general, including a discussion of initial values and stopping rules. Then we define expected choice probabilities and reaction times and state that these exist and can be calculated in finite time. The proof of this statement is given later in the paper. It is not essential for understanding the theory; we provide it rather as completing the theoretical derivations. Next, we show how to derive transition probabilities from the given alternatives in a specific choice set and therewith define the random walk for that set. A psychological interpretation of their constituents is given afterward. Finally, we demonstrate the predictive power of our model by showing several simulations for choice situations producing the similarity, attraction, and compromise effect and calculate expected hitting times and choice probabilities. We conclude with a comparison of the $2N$-ary choice tree with its closest competitors, the multi-alternative DFT and the LCA model.

2. THE $2N$-ARY CHOICE TREE MODEL

Making an informed decision usually implies sampling of information about the alternatives under consideration. In Psychology, information sampling processes (e.g., Townsend and Ashby, 1983; Luce, 1986, for review, LaBerge, 1962; Laming, 1968; Link and Heath, 1975; Townsend and Ashby, 1983; Luce, 1986; Ratcliff and Smith, 2004) have a long tradition and proven to be an adequate tool for detailed interpretation of decision making processes, mostly in perception as they provide insight about accuracy and time course of these processes. Poisson counter models (e.g., Pike, 1966; Townsend and Ashby, 1983; LaBerge, 1994; Diederich, 1995; Smith and Van Zandt, 2000; Van Zandt et al., 2000) are a special class of information sampling models that assume constant amounts of information being sampled at Poisson distributed points in time. (Multi-alternative) DFT (Busemeyer and Townsend, 1993; Roe et al., 2001) and the LCA model (Usher and McClelland, 2004) make use of information sampling principles in modeling preferential choice under uncertainty. Both models assume one counter per alternative and all of these counters are updated once per fixed time interval until one of them reaches a threshold. The amounts to update the counters depend

on comparison of the alternatives and on already sampled information. In our $2N$-ary choice tree model, only one counter per fixed time interval is updated with a fixed amount, but the probability for each counter to be updated depends on comparison of the alternatives and on already sampled information. With regard to its constituents it is thus based on the same principles as both DFT and the LCA model. As only one counter is updated per iteration, the next time for a specific counter to be updated depends on the given probabilities. Hence the technical component of the $2N$-ary choice tree model resembles a counter model.

2.1. $2N$-ARY CHOICE TREES

In contrast to the aforementioned models, the $2N$-ary choice tree model assigns two counters to each of N alternatives in a given choice set. One of them samples positive information, i.e., information in favor of the respective alternative, the other one samples negative information, i.e., information against it. Their difference describes the actual preference state relating to that alternative. As an example, consider two alternatives A and B. The four counters are labeled A^+, A^-, B^+, and B^- and yield the preference states $Pref(A) = A^+ - A^-$ for alternative A and $Pref(B) = B^+ - B^-$ for alternative B. Beginning at a fixed point in time, the model chooses one counter and increases its state by one whenever a specific time interval h (e.g., 1 ms) has passed. The length h of the time interval can be chosen arbitrarily with a shorter time interval leading to more precision in the calculation of expected choice probabilities and choice response times. Due to limitations of recording devices, experimental data will be discrete as well and it is thus not necessary to aim for a continuous model. Note that increasing only one counter state at a time with a fixed amount of evidence equal to one is equivalent to increasing all counter states at the same time with an amount of evidence equal to the probability with which these counters are chosen and which also sum up to one (see below). Updating counters at discrete points in time creates a discrete structure of possible combinations of counter states which can be interpreted as graph or, more precisely, as (b-ary) tree[1].

Definition 1 (b-ary tree): A b-ary tree is a rooted tree $T = (V, E, r)$ with vertices V, edges $E \subseteq V \times V$ and root $r \in V$ where all vertices are directed away from r and each internal vertex has b children.

For N choice alternatives, consider a $2N$-ary tree $T = (V, E, r)$. **Figure 1** depicts the 4-ary tree for the two-alternative example. The topmost vertex is the root r with outgoing edges directing to four vertices that represent the counters A^+, A^-, B^+, and B^-. Each of these vertices has four outgoing edges and thus four children itself, and so forth. The information sampling process is mapped to this tree as a walk, i.e., a sequence of edges and vertices, beginning with the root r that takes one step, i.e., passes from one vertex through an edge to another vertex, per time interval h. Whenever the walk reaches a vertex, the counter with the same label is updated by $+1$. **Figure 2** shows an example for a walk on the 4-ary tree where first counter B^+ (information in favor of choice alternative B), then counter A^- (unfavorable information for choosing alternative A) and then again counter B^+ is updated.

---

[1]Definitions of graph-related terms not defined here can be found in Korte and Vygen (2002).
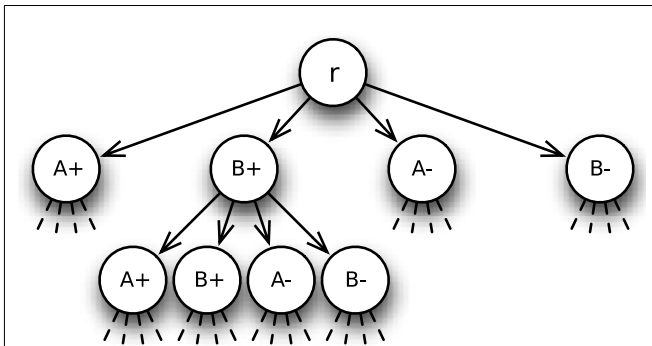
FIGURE 1 | 4-ary tree for choice between two alternatives *A* and *B*. The root *r* has four outgoing edges directing to four vertices that represent the counters A^+, A^-, B^+, and B^-. Each of these vertices has four outgoing edges and thus four children itself, and so forth.
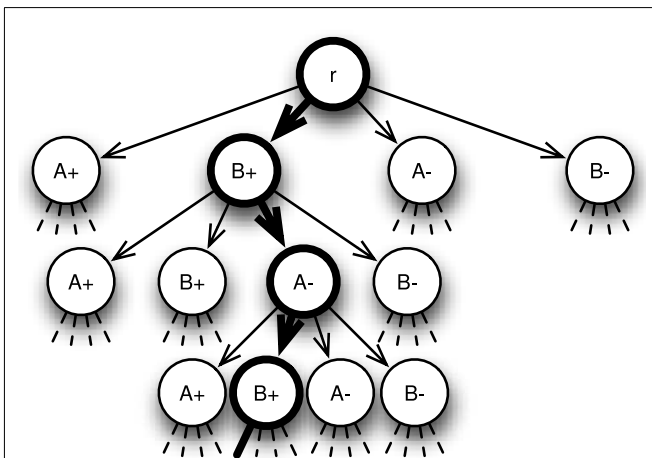

FIGURE 3 | Transition probabilities for the two-alternative choice problem with no counter-dependent and random component. $p(A^+)$, $p(B)^+$, $p(A)^+$, and $p(B^-)$ sum up to one and are the same for the outgoing edges of each vertex $v \in V$.

FIGURE 2 | 4-ary tree for choice between two alternatives *A* and *B* with highlighted sample path $r \to B^+ \to A^- \to B^+ \to \ldots$

Three features of the model are of specific interest: (a) when and how the walk starts after presentation of choice alternatives (in an experimental trial), (b) how the walk chooses the next edge to pass through in each step, and (c) when and how the walk stops. Without an *a priori* bias toward any of the choice alternatives, we assume that all counter states are set to zero at the outset of the information sampling process and hence, the process starts with presentation of the choice alternatives. Biases toward one or several of the alternatives can be implemented by either defining initial values unequal to zero for these alternatives or by independently sampling information for the alternatives from predefined distributions for some time before the actual information sampling process starts (cf. Diederich and Busemeyer, 2006; Diederich, 2008). For simplicity, we assume no biases here, i.e., initial values are set to zero for all alternatives. Note that for the 2*N*-ary choice tree, initial values are counter states at the root *r*. Then the walk moves away from there step by step, choosing the next edge to pass through by means of so-called transition probabilities p_e, $e \in E$. The transition probabilities are built up of
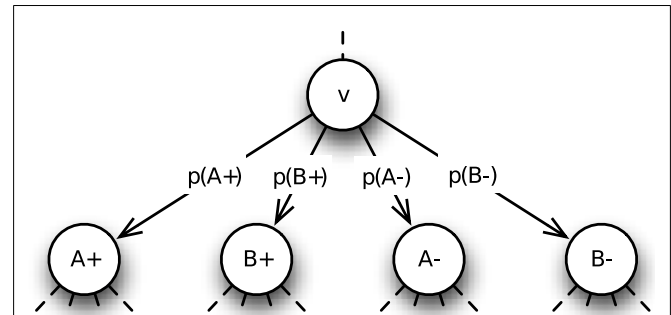
the comparison of the alternatives the decision maker considers and supplemented with a counter-dependent component and a random component. For each vertex, the transition probabilities for all outgoing edges sum up to one, so that the walk does not stay still at any vertex it reaches throughout the information sampling process. We show the structure of the model first; a detailed description of the transition probabilities is presented in the next section. For simplicity consider a choice situation with two alternatives *A* and *B*; the counter-dependent component and the random component are set to zero. As shown in **Figure 3**, transition probabilities are the same for the outgoing edges of each vertex $v \in V$, i.e., $p_{(v,v(A^+))} = p(A^+)$ for each edge $(v,v(A^+)) \in E$ leading to a vertex with label A^+, $p_{(v,v(B^+))} = p(B^+)$ for each edge leading to a vertex with label B^+ and so on for the other counters A^- and B^-. The probability for walking along a specific path is the product of transition probabilities of all edges on that path. In our example, the probability p for making the first three steps as shown in **Figure 2** is $p = p(B^+) \cdot p(A^-) \cdot p(B^+)$.

The third topic addresses the stopping rule, that is, when the decision maker stops sampling information and chooses a choice alternative. A specific stopping rule depends on the preference states associated with the alternatives, i.e., the differences of their respective two counters which are compared to certain thresholds θ. The thresholds can be defined in several ways, their suitability depending on the definition of transition probabilities and initial values. They are (1) one single positive threshold $\theta^+ > 0$ for all alternatives, (2) one positive and one negative threshold $\theta^+ > 0$ and $\theta^- < 0$ for all alternatives, (3) a positive threshold θ_i^+ for each alternative $i \in \{1, \ldots, N\}$, and (4) a positive and a negative threshold θ_i^+ and θ_i^- for each alternative $i \in \{1, \ldots, N\}$.

Obviously, the simplest setup is a single positive threshold $\theta^+ > 0$ for all alternatives, which is hit as soon as the information sampled in favor of any/one of the alternatives exceeds the information against it by θ^+ for the first time, i.e., when $Pref(i) = \theta^+$ for one alternative $i \in \{1, \ldots, N\}$. Sometimes, however, the probability for collecting negative information may be greater than the probability for sampling information in favor of these alternatives and reaching a positive threshold θ^+ is very unlikely. For those situations it is useful to introduce a second, negative threshold, $\theta^- < 0$, which is hit when negative information of one alternative

exceeds the positive information of this alternative by $-\theta^-$, i.e., $\mathit{Pref}(i) = \theta^-$. In this case the respective alternative is not chosen but withdrawn from the choice set and the sampling process continues with one alternative less as described in the next paragraph. Note, that in both cases the thresholds are global in the sense that the same thresholds apply for all choice alternatives. Finally, θ^+ (and θ^-) may vary from alternative to alternative, yielding one (or two) thresholds θ_i^+ (and θ_i^-) for each of N alternatives, $i \in \{1, \ldots, N\}$. Here the thresholds are local in the sense that each alternative has its own threshold(s). This is an alternative way to implement biases when the initial values are zero. That is, biases do not affect transition probabilities through the counter-dependent component and can thus be interpreted as the decision maker's stable opinion about the presented alternatives.

Withdrawal of alternatives from a choice set traces back to the model of elimination by aspects (EBA model, Tversky, 1972). But whereas elimination is the only means to come to a decision in the EBA model, the $2N$-ary choice tree model like the multi-alternative DFT (Roe et al., 2001) provides several ways to reach a decision. An alternative i is chosen either if its preference state exceeds θ_i^+ or if all other alternatives have been withdrawn from the choice set or a combination of these two. **Figure 4** shows three examples of walks that lead to the choice of alternative B from a set of three alternatives A, B, and C with global thresholds $\theta^+ = 2$ and $\theta^- = -2$. The leftmost walk represents direct choice of alternative B, the rightmost withdrawal of alternative C and subsequent choice of option B and the middle walk illustrates withdrawal of alternative A first and then of option C. Note that after withdrawal of one alternative, there are two-outgoing edges less from the respective vertex downward. Transition probabilities change accordingly, i.e., the withdrawn alternative is removed from the comparison procedure and its counter states no longer contribute to the counter-dependent component (cf. next section). This corresponds to an anew started information sampling process between the remaining alternatives and their previous counter states as initial values.

For a choice set with N alternatives and given thresholds $\theta_i^\pm$ this defines the structure of the $2N$-ary choice tree. For each alternative $i \in \{1, 2, \ldots, N\}$ we can thus completely identify the set $V_i \subseteq V$ of vertices where alternative i is chosen. Defining $P_v := P_{\{r, v\}}$ to be the unique path from the root r to a vertex $v \in V$ and given transition probabilities p_e for all edges $e \in E$ we can identify the probability for walking along a path P_v as the product $p_v = \Pi_{e \in P_v} p_e$ and therewith define:

Definition 2 (expected choice probability): The expected probability for choosing alternative $i \in \{1, 2, \ldots, N\}$ is the probability for reaching the set V_i:

$$p_i = \sum_{v \in V_i} \prod_{e \in P_v} p_e. \tag{1}$$

The length $|P_v|$ of the path P_v from r to $v \in V$ indicates the number of steps that the random walk has to take to reach v. Multiplied by the length h of the time interval, this yields the time it takes to cover the distance from r to v. Thus $T_v = h \cdot |P_v|$.

Definition 3 (expected hitting time): The expected time for choosing alternative $i \in \{1, 2, \ldots, N\}$ is the sum of expected hitting
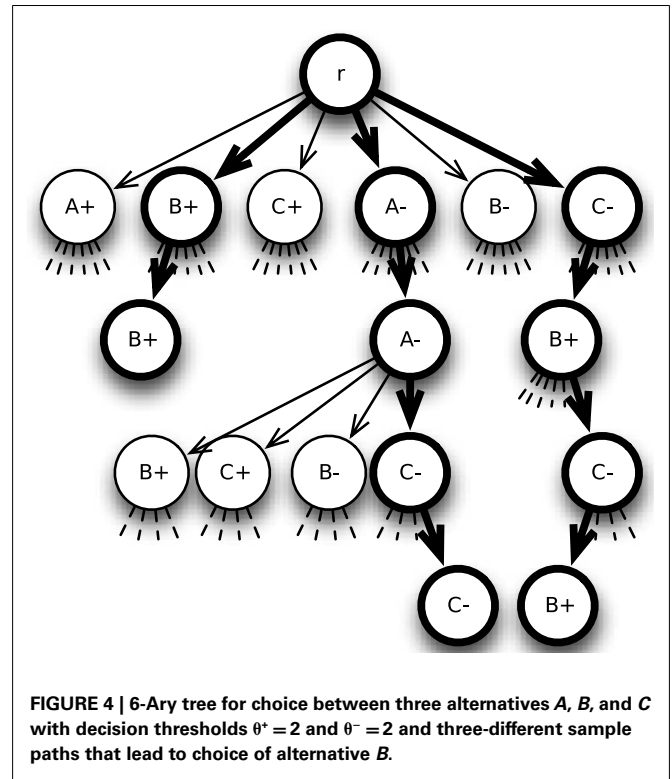

FIGURE 4 | 6-Ary tree for choice between three alternatives *A*, *B*, and *C* with decision thresholds $\theta^+ = 2$ and $\theta^- = 2$ and three-different sample paths that lead to choice of alternative *B*.

times for each vertex $v \in V_i$ weighted by the probability for reaching v:

$$\mathbb{E}[T_i] = h \cdot \sum_{v \in V_i} |P_v| \cdot p_v = h \cdot \sum_{v \in V_i} |P_v| \cdot \prod_{e \in P_v} p_e. \tag{2}$$

The expected choice probabilities and hitting times can be approximated up to absolute accuracy in finite time. See below for the formal statement and proof of this property.

2.2. TRANSITION PROBABILITIES

Having defined the skeletal structure of our theory, we can now proceed to its heart, the transition probabilities. The main components are (a) weighted comparison of alternatives, (b) mutual or global inhibition, (c) decay of already sampled information over time, and (d) a random part. The transition probabilities control the information sampling process and thus describe the development of human preferences in specific choice situations. Throughout this section we will consider such situations with N choice alternatives that are evaluated with respect to the same D attributes. For each alternative, the decision maker is provided with one non-negative value per attribute, representing an objective evaluation of that alternative with respect to the attributes. The $N \cdot D$ values in total can be stored in a $N \times D$-matrix $L = (l_{ij})$ with $i = 1, \ldots, N$ and $j = 1, \ldots, D$.

The definition of transition probabilities is based on weighted integration of results of an attribute-wise comparison of alternatives. To ensure equally significant impact of the weight parameters, preprocessing of the values of the alternatives with respect to the attributes is necessary and we do so by separately normalizing them to one for each attribute. This yields a new

matrix $M = (m_{ij}) = (l_{ij}/\sum_{k=1}^{N} l_{kj})$ with $i=1, \ldots, N$ and $j=1, \ldots, D$ and thus column sum $\sum_{i=1}^{N} m_{ij} = 1$ for all $j \in \{1, \ldots, D\}$.

2.2.1. Comparison of alternatives

At first we focus on one attribute j only. The easiest way to define transition probabilities is simply to use the entries from the respective column j of M and assign them to the edges that affiliate with the counters for positive information. The counters for negative information get transition probabilities zero. This corresponds to a framework where the alternatives are compared to an inferior external reference point (e.g., cars A, B, and C are compared to not having a car at all). Because the values for each attribute sum up to one already, no further normalization is needed.

We differentiate between external reference points that are not part of the choice set and internal reference points that are part of the choice set and available for the decision maker. For instance, if someone moves to a new city and has to choose between several available apartments, she will probably compare them to her old apartment which is no longer available in the new city and thus an example for an external reference point. Or consider a choice of dessert in a restaurant when the decision maker is told that the chocolate cake she ordered is no longer available because someone just had the last piece. An internal reference point, however, is part of the choice set, actually several or even all available alternatives can be used as internal reference points at the same time, possibly in combination with an external reference point.

Having decided which reference points to use, the alternatives $i \in \{1, \ldots, N\}$ are compared to them. For each alternative i, favorable and unfavorable comparisons are handled separately and the absolute values of their differences are summed up to obtain measures of evidence for and against alternative i, respectively. This yields two non-negative values per alternative and thus $2N$ values in total that are then normalized to one in order to obtain probabilities. In the car example where three cars are compared to not having a car at all, probabilities associated with negative counters are set to zero as each car is better than, presumably, no car. Actually, whenever one single reference point is used, at least half of the probabilities are zero because each alternative is either favored over the reference point or not and hence there cannot be evidence for and against one alternative at the same time. However, our main focus is on situations where each of at least three alternatives in the choice set is used as internal reference point for all the other alternatives and thus there are at least two-reference points.

In this case we obtain a vector

$$P_j = \begin{pmatrix} p_{1j} \\ \vdots \\ p_{Nj} \\ p_{(N+1)j} \\ \vdots \\ p_{(2N)j} \end{pmatrix} = \begin{pmatrix} p_{1j}^{+} \\ \vdots \\ p_{Nj}^{+} \\ p_{1j}^{-} \\ \vdots \\ p_{Nj}^{-} \end{pmatrix} / \sum_{i=1}^{N} \left(p_{ij}^{+} + p_{ij}^{-} \right)$$

with

$$p_{ij}^{+} = \sum_{k \neq i} \left(m_{ij} - m_{kj} \right) \cdot \mathbb{I} \left(m_{ij} > m_{kj} \right),$$

$$p_{ij}^{-} = \sum_{k \neq i} \left(m_{kj} - m_{ij} \right) \cdot \mathbb{I} \left(m_{ij} < m_{kj} \right)$$

for $k = 1, \ldots N$ and

$$\mathbb{I}(x) = \begin{cases} 1, & \text{if } x \text{ is true} \\ 0, & \text{else.} \end{cases}$$

Especially with an external reference point at hand, the actual choice may lead to a loss of some kind. For instance, in the apartment example above a loss could be a further way to the workplace or a smaller bathroom. People usually try to avoid losses more than they seek gains while overrating small losses and gains compared to larger ones (Kahneman and Tversky, 1991; Tversky and Simonson, 1993). The 2*N*-ary choice tree model can account for the loss-aversion principle (Kahneman and Tversky, 1979) with an asymmetric value function (Kahneman and Tversky, 1979; Tversky and Simonson, 1993) by increasing probabilities for sampling negative information compared to probabilities for gathering positive information.

In their LCA model, Usher and McClelland (2004) use an asymmetric value function

$$V(x) = \begin{cases} \log(1 + x), & \text{for } x > 0 \\ -\left[\log\left(1 + |x|\right) + \log\left(1 + |x|\right)^2 \right], & \text{for } x < 0 \end{cases}$$

and apply it to the relative advantages ($x > 0$) and disadvantages ($x < 0$) of alternatives compared to each other on one dimension. $V(x)$ is steeper for losses than for gains but flattens for both advantages and disadvantages when they become bigger. This favors similar pairs of alternatives over dissimilar ones and allows the LCA model to account for attraction and compromise effects.

Adopting it to our 2*N*-ary choice tree model this yields an asymmetric value function

$$A(x) = \begin{cases} \log(1 + x), & \text{for favorable comparisons} \\ \log(1 + x) \\ \quad + \log\left(1 + x\right)^2, & \text{for unfavorable comparisons,} \end{cases}$$

which can be applied to the absolute differences from the comparison process before normalizing them to one:

$$p_{ij}^{+} = \sum_{k \neq i} A\left(m_{ij} - m_{kj} \right) \cdot \mathbb{I} \left(m_{ij} > m_{kj} \right),$$

$$p_{ij}^{-} = \sum_{k \neq i} A\left(m_{kj} - m_{ij} \right) \cdot \mathbb{I} \left(m_{ij} < m_{kj} \right).$$

The asymmetric value function $A(x)$ is not necessary for explanation of similarity, attraction, or compromise effects in the 2*N*-ary choice tree model but moderates the strength of the compromise effect (see below). In cases where θ^{+} and θ^{-} have the same

order of magnitude, application of $A(x)$ leads to faster withdrawal of alternatives and hence, more decisions are based on withdrawal of all but one alternative.

In summary, the comparison of alternatives provides us with a set of $2N$ transition probabilities for each attribute $j \in \{1, \ldots, D\}$ that form a vector P_j. Each of these vectors can be used to model an information sampling process based on a single attribute. As the probabilities are derived from comparison of alternatives only, they remain constant during the whole process.

2.2.2. Weighting of attributes

So far we have only focused on one attribute but choice alternatives in real life are most often described by several attributes and thus require more elaboration. In the following, we consider choice sets with N alternatives characterized by $D \geq 2$ attributes. Especially situations with three alternatives where similarity, attraction or compromise effects have been observed, require at least two attributes to distinguish the different alternatives from each other. Note that it is difficult to construct a choice set with two equally attractive but different alternatives due to the decision maker's individual salience. Diederich (1997) accounts for subjective salience by defining a Markov process on the attributes giving probabilities for switching attention from one attribute to the other. This process can be directly implemented into the transition probabilities by using a stationary distribution on the attributes. Each attribute $j \in \{1, \ldots, D\}$ is assigned a weight w_j that corresponds to the probability for considering this attribute during the information sampling process. For each alternative $i \in \{1, \ldots, N\}$ weighted positive and negative evidence is added up and normalized to obtain a proper probability distribution (the probabilities add up to one), that is

$$P = \begin{pmatrix} p_1 \\ \vdots \\ p_N \\ p_{(N+1)} \\ \vdots \\ p_{(2N)} \end{pmatrix} = \begin{pmatrix} p_1^+ \\ \vdots \\ p_N^+ \\ p_1^- \\ \vdots \\ p_N^- \end{pmatrix} / \sum_{i=1}^{N} \left(p_i^+ + p_i^- \right)$$

with $p_i^+ = \sum_{j=1}^{D}(p_{ij}^+ \cdot w_j)$ and $p_i^- = \sum_{j=1}^{D}(p_{ij}^- \cdot w_j)$ for $i \in \{1, \ldots, N\}$.

The weights account for subjective salience that in turn may be influenced by several internal and external factors such as personal preferences, social influences, characteristics of the choice set, or the experimenter's instructions. Personal preferences like, for instance, the preference of time over money or of tastiness to healthiness may be learned from friends, family, or other people in our surrounding. They are generally independent from the choice situation and hence, their impact on the information sampling is indirect. On the contrary, the choice set itself has a direct influence on the subjective saliences. For example, the decision maker may primarily focus on those attributes where alternatives are very similar to each other, because this information may be crucial for the choice. Or she concentrates on attributes with somehow outstanding values. It is therefore important to normalize the values for each attribute as described before because this guarantees

representation of these effects by the attention weights. People that are present during the deliberation process like sales people or immediately prior to it like the experimenter in a laboratory context can also have a direct influence on the saliences by drawing the decision maker's attention to a specific attribute. This can be used to verify influence of attention weights by instructing decision makers to focus on certain attributes while choosing between different cars, salad dressings, chocolate bars or shoes. Corresponding experiments are under way.

2.2.3. Noise

In order to account for random fluctuations in the decision maker's attention (cf. Busemeyer and Townsend, 1993) which are independent of the characteristics of the choice alternatives, we add a constant to each transition probability. This makes every outgoing edge of a vertex $v \in V$ available for the next (random) step because it guarantees non-zero transition probabilities for all of them. Let $\mathcal{N}$ be a vector of length $2N$ with all entries equal to $1/2N$. Weighting the transition probabilities P from the weighted comparison of alternatives by $(1 - \xi)$ with $0 \leq \xi \leq 1$ and adding the product $\xi \cdot \mathcal{N}$ yields noisy transition probabilities where ξ moderates the strength of the uniformly distributed noise:

$$P_{\mathcal{N}} = (1 - \xi) \cdot P + \xi \cdot \mathcal{N}.$$

The vector $P_{\mathcal{N}}$ of noisy transition probabilities integrates comparison of alternatives on all present attributes. Related to the $2N$-ary choice tree, this information is global as it is independent of the local counter states and thus the transition probabilities are the same for the edges emanating from each vertex.

2.2.4. Leakage

During their development of DFT (Busemeyer and Townsend, 1993) introduce a factor s for serial positioning effects, called "growth-decay rate." It produces recency effects for $0 < s < 1$ and primacy effects for $s < 0$. In their multi-alternative version of DFT (Roe et al., 2001) the reverse $(1 - s)$ of this factor reappears as "self-feedback loop" and accounts for the memory of previous preference states. $(1 - s) = 1$ denotes perfect memory of the previous state $(1 - s) = 0$ no memory at all. For their simulations (Roe et al., 2001) use $(1 - s) = 0.94$ or $(1 - s) = 0.95$. Usher and McClelland (2001, 2004) adopted the idea of the self-feedback loop, but call it "leakage" λ and – based on findings from neuroscience – interpret it as "neural decay."

In order to account for decay of already sampled information over time, we implement leakage $\mathcal{L}$ into our transition probabilities. Leakage obviously depends on already sampled information and thus we normalize the current states of our $2N$ counters to $1 - \lambda$ and for each alternative $i \in \{1, \ldots, N\}$ add the result for positive (negative) counter of alternative i to the transition probability associated with the negative (positive) counter for alternative i weighted by λ. Like this, the overall sum of the transition probabilities remains 1 and only $100 \cdot \lambda\%$ of the sampled information is actually memorized. The greater λ, the longer it takes until the process reaches a threshold. Overall, this yields

$$P_{\mathcal{N}\mathcal{L}} = (1 - \lambda) \cdot [(1 - \xi) \cdot P + \xi \cdot \mathcal{N}] + \lambda \cdot \mathcal{L}.$$

2.2.5. Inhibition

To account for the similarity, attraction, and compromise effect, DFT (Roe et al., 2001) and the LCA model (Usher and McClelland, 2004) both rely on inhibition. Whereas distance-dependent inhibition enables DFT to account for the attraction and compromise effect, global inhibition produces the similarity effect in the LCA model. We can implement both types of inhibition into the 2N-ary choice tree model to explore their impact on the aforementioned effects. We define weights for all pairs of alternatives by either using the same weight for all pairs like Usher and McClelland (2004) do with their "global inhibition" parameter β or different weights like Roe et al. (2001) do with their distance-dependent weights (i.e., higher weights for more similar alternatives). Those weights can be stored in a symmetric $N \times N$-matrix with zeros on the diagonal.

Taking into account the basic concept of inhibition, we assume that the state of the positive counter for each alternative $i \in \{1, \ldots, N\}$ reduces sampling of positive information for all other alternatives $j \in \{1, \ldots, N\} - \{i\}$. Because this is equivalent to increasing sampling of negative information for these alternatives and vice versa for states of negative counters, we implement inhibition $\mathscr{I}$ into our model as follows: Multiplying the symmetric $N \times N$-matrix with both the vector of states of positive counters and negative counters yields two vectors with weighted sums of counter states. We concatenate them in inverted order and normalize the resulting vector of length $2N$ to μ before adding it to the vector of transition probabilities now weighted by $(1 - \lambda - \mu)$. This completes the final definition of transition probabilities

$$P_{\mathscr{N}\mathscr{L}\mathscr{I}} = (1 - \lambda - \mu) \cdot [(1 - \xi) \cdot P + \xi \cdot \mathscr{N}] + \lambda \cdot \mathscr{L} + \mu \cdot \mathscr{I}.$$

In a nutshell, the transition probabilities consist of a global part that is independent from the current counter states of the random walk and a local part that depends on already sampled information. The global parts are weighted sums of comparative values P and noise $\mathscr{N}$ that remain constant during the whole process. They are complemented with leakage $\mathscr{L}$ and inhibition $\mathscr{I}$ which may change from step to step and hence, are local in the terminology of the 2N-ary choice tree model.

3. PREDICTIONS OF THE 2N-ARY CHOICE TREE MODEL

To show the predictions of the 2N-ary choice tree model and how it accounts for similarity, attraction, and compromise effects in choice settings with three alternatives characterized by two attributes, we run several simulations. An extension to more alternatives is straightforward. We will define values l_{ij} that range between 0 and 10. As values of choice alternatives are normalized to one on each dimension before comparison, only the relative amount of these values is of importance. Unless stated otherwise, we run 1000 trials per simulation with threshold $\theta = 20$, noise factor $\xi = 0.01$ and leakage factor $\lambda = 0.05$, but without inhibition (i.e., $\mu = 0$).

In order to meet the assumptions of the similarity, attraction, and compromise effect, we constructed two equally attractive but dissimilar alternatives $A = (9, 1)$ and $B = (1, 9)$ that are both evaluated with respect to two attributes. The choice probabilities were 0.52, 0.51, and 0.47 for alternative A and 0.48, 0.49, and 0.53 for option B in three simulations with the above mentioned parameters and attribute weight 0.5 for both attributes.

To reproduce the similarity effect (Simonson, 1989) we add a third alternative to the choice set that is equal or similar to either option A or B, i.e., $C = (1, 9)$, $C_2 = (0.9, 9.1)$, or $C_3 = (1.1, 8.9)$. To prevent a combination of the similarity effect with a slight compromise effect (cf. Usher and McClelland, 2004), we will use only C for demonstration, but the results for options C_2 and C_3 are very similar to the ones presented here. The alternatives are put together in a 3×2-matrix L, whose columns are normalized to one, resulting in matrix M:

$$L = \begin{pmatrix} 9 & 1 \\ 1 & 9 \\ 1 & 9 \end{pmatrix} \quad \text{and} \quad M = \begin{pmatrix} 0.818 & 0.053 \\ 0.091 & 0.474 \\ 0.091 & 0.474 \end{pmatrix}.$$

M already shows smaller values for alternatives B and C on the second dimension than for alternative A on dimension one which characterizes the similarity effect. In the next step, the values on each dimension are compared to each other, resulting in a 6×2-matrix that is then multiplied by $W = \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix}$ before being normalized to one again:

$$P' = \begin{pmatrix} 1.4545 & 0 \\ 0 & 0.4211 \\ 0 & 0.4211 \\ 0 & 0.8421 \\ 0.7273 & 0 \\ 0.7273 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix} = \begin{pmatrix} 0.5818 \\ 0.2526 \\ 0.2526 \\ 0.5053 \\ 0.2909 \\ 0.2909 \end{pmatrix}.$$

and

$$P = \begin{pmatrix} 0.5818 \\ 0.2526 \\ 0.2526 \\ 0.5053 \\ 0.2909 \\ 0.2909 \end{pmatrix} / 2.1742 = \begin{pmatrix} 0.2676 \\ 0.1162 \\ 0.1162 \\ 0.2324 \\ 0.1338 \\ 0.1338 \end{pmatrix}.$$

Finally noise is added to this constant part of the transition probabilities. In contrast to leakage that depends on the respective counter states and has to be computed anew for every step, $P_{\mathscr{N}}$ remains constant over time. The only occasion where it changes is after withdrawal of one alternative from the choice set.

The most interesting parameters in this attempt to model a similarity effect are the attribute weights as they control the strength of the effect. **Figure 5** demonstrates this by means of choice probabilities from simulations with different sets of attribute weights but otherwise unchanged parameters. It starts with $W = \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix}$ and $W = \begin{pmatrix} 0.55 \\ 0.45 \end{pmatrix}$ on the left side and gradually changes by 0.05 to $W = \begin{pmatrix} 0.25 \\ 0.75 \end{pmatrix}$ on the right side. The relative frequency of choices for alternatives A, B, and C including the mean number of steps leading to these choices can be found in **Table 1**.

The same mechanisms account for the attraction effect (Huber et al., 1982) that occurs during choice between two equally attractive but dissimilar alternatives A and B and a third alternative C

**FIGURE 5 | Choice probabilities for choice between three alternatives
$A = (9, 1)$, $B = (1, 9)$, and $C = (1, 9)$ and different attention weights w_1
and w_2 for the two attributes.** The abscissa is labeled with increasing
values of w_2 corresponding to decreasing values of w_1. For $w_2 < 0.625$ a
similarity effect can be observed.

**FIGURE 6 | Choice probabilities for choice between three alternatives
$A = (9, 1)$, $B = (1, 9)$, and $C = (1, 8.5)$ and different attention weights w_1
and w_2 for the two attributes.** The abscissa is labeled with increasing
values of w_2 corresponding to decreasing values of w_1. For
$0.61 \leq w_2 < 0.65$ an attraction effect can be observed.
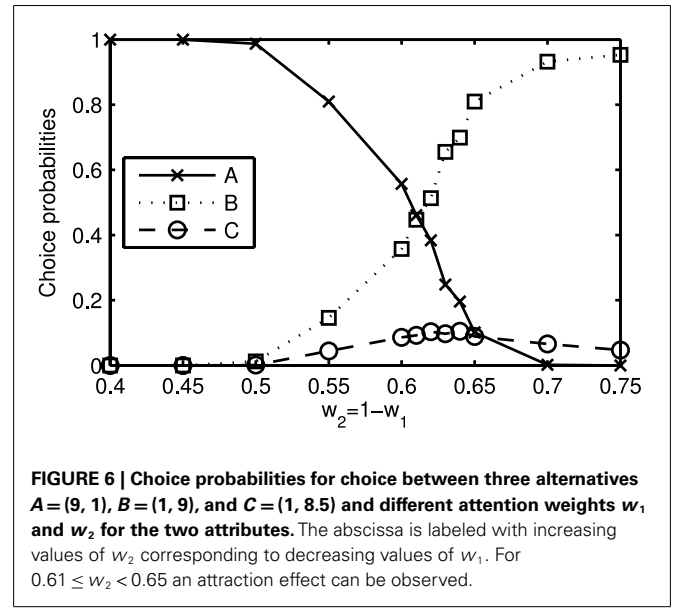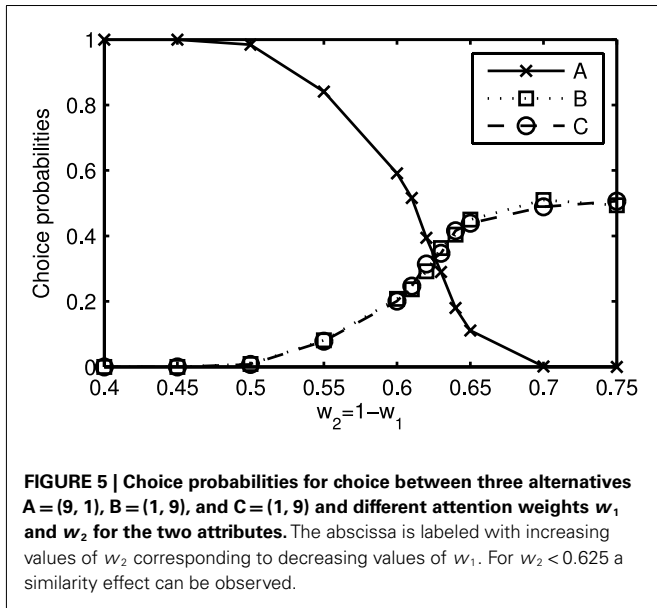
**Table 1 | Relative number of choices and mean response times
(arbitrary unit, in parentheses) for alternatives $A = (9, 1)$, $B = (1, 9)$, and
$C = (1, 9)$ from 1000 simulations with $\theta = 20$, $\xi = 0.01$, $\lambda = 0.05$, $\mu = 0$,
and $w_2 = 1 - w_1$ ranging from 0.4 to 0.75 as indicated in the first row.**

w_2	0.4	0.45	0.5	0.55
A	1 (92.4)	1 (116.4)	0.984 (163.0)	0.841 (205.5)
B	0 (–)	0 (–)	0.009 (596.4)	0.081 (555.6)
C	0 (–)	0 (–)	0.007 (475.4)	0.078 (524.6)
w_2	0.6	0.61	0.62	0.63
A	0.591 (316.5)	0.516 (358.3)	0.394 (401.0)	0.29 (383.7)
B	0.208 (554.7)	0.237 (622.1)	0.292 (602.8)	0.363 (666.6)
C	0.201 (568.9)	0.247 (590.1)	0.314 (634.0)	0.347 (666.9)
w_2	0.64	0.65	0.7	0.75
A	0.180 (418.9)	0.111 (430.7)	0.001 (124.0)	0 (–)
B	0.404 (685.3)	0.450 (679.4)	0.510 (631.8)	0.494 (658.5)
C	0.416 (689.4)	0.439 (678.6)	0.489 (664.2)	0.506 (666.1)

**Table 2 | Relative number of choices and mean response times
(arbitrary unit, in parentheses) for alternatives $A = (9, 1)$, $B = (1, 9)$, and
$C = (1, 8.5)$ from 1000 simulations with $\theta = 20$, $\xi = 0.01$, $\lambda = 0.05$, $\mu = 0$,
and $w_2 = 1 - w_1$ ranging from 0.4 to 0.75 as indicated in the first row.**

w_2	0.4	0.45	0.5	0.55
A	1 (93.5)	1 (118.4)	0.987 (168.7)	0.81 (204.7)
B	0 (–)	0 (–)	0.012 (593.3)	0.146 (523.4)
C	0 (–)	0 (–)	0.001 (499.0)	0.044 (601.7)
w_2	0.6	0.61	0.62	0.63
A	0.557 (313.1)	0.461 (343.6)	0.384 (391.2)	0.248 (396.5)
B	0.357 (466.2)	0.447 (476.1)	0.513 (473.6)	0.655 (487.3)
C	0.086 (558.7)	0.092 (568.1)	0.103 (569.2)	0.097 (605.3)
w_2	0.64	0.65	0.7	0.75
A	0.196 (395.5)	0.101 (342.5)	0.002 (155.5)	0 (–)
B	0.699 (448.9)	0.810 (414.6)	0.932 (237.2)	0.953 (159.5)
C	0.105 (483.7)	0.089 (459.0)	0.066 (243.7)	0.047 (174.6)

that is similar to one of these but slightly less attractive. For $A = (9, 1)$, $B = (1, 9)$, and $C = (1, 8.5)$ this yields

$$L = \begin{pmatrix} 9 & 1 \\ 1 & 9 \\ 1 & 8.5 \end{pmatrix} \quad \text{and} \quad M = \begin{pmatrix} 0.818 & 0.054 \\ 0.091 & 0.487 \\ 0.091 & 0.459 \end{pmatrix}.$$

As shown in **Figure 6**, the attraction effect occurs between $W = \begin{pmatrix} 0.39 \\ 0.61 \end{pmatrix}$ and $W = \begin{pmatrix} 0.35 \\ 0.65 \end{pmatrix}$ Note that the deviation from weights $w_1 = w_2 = 0.5$ is due to a higher salience of attribute two because the values on this attribute differentiate between the alternatives. The relative frequency of choices for alternatives A, B, and C including the mean number of steps leading to these choices can be found in **Table 2**.

For the compromise effect (Simonson, 1989), two equally attractive but dissimilar alternatives $A = (9, 1)$ and $B = (1, 9)$

compete against a compromise option $C = (5, 5)$. Note that the defined values for each alternative sum up to ten and thus all three alternatives objectively are equally attractive provided the attributes are equally weighted. We get

$$L = \begin{pmatrix} 9 & 1 \\ 1 & 9 \\ 5 & 5 \end{pmatrix} \quad \text{and} \quad M = \begin{pmatrix} 0.6 & 0.067 \\ 0.067 & 0.6 \\ 0.333 & 0.333 \end{pmatrix},$$

and restricting $w_1 = w_2 = 0.5$ to be equal, this yields

$$P = \begin{pmatrix} 0.4 \\ 0.4 \\ 0.2667 \\ 0.4 \\ 0.4 \\ 0.2667 \end{pmatrix}.$$

So far, these transition probabilities do not seem to induce any compromise effect but as the probabilities for sampling negative information are comparatively high, withdrawal of one alternative from the choice set frequently occurs in that setting. After withdrawal of one alternative, comparison of the remaining alternatives is renewed. In the cases where alternative A or B are withdrawn, the new probabilities clearly favor the compromise option C, yielding an overall preference for that alternative:

$$M_{-A/-B} = \begin{pmatrix} 0.643 & 0.167 \\ 0.357 & 0.833 \end{pmatrix}, \; P_{-A/-B} = \begin{pmatrix} 0.15 \\ 0.35 \\ 0.35 \\ 0.15 \end{pmatrix}.$$

In 1000 trials with decision threshold $\theta = 20$, noise factor $\xi = 0.01$, leakage factor $\lambda = 0.05$, and no inhibition, alternatives A and B were chosen 247 (24.7%) and 250 (25%) times, respectively, and option C won 503 (50.3%) decisions. Decreasing θ to 10 yields choice frequencies of 243 (24.3%) for alternative A, 267 (26.7%) for option B, and 490 (49%) for alternative C. $\theta = 5$ leads to 253 (36.9%) choices with an average step number of 36.9 for alternative A, 269 (26.9%) choices with 38.3 steps on average for option B, and 478 (47.8%) choices with 43.8 steps on average for alternative C. **Figure 7** shows the response time distribution for alternative A for $\theta = 5$, $\xi = 0.01$, $\lambda = 0.05$, and $\mu = 0$. The expected response time, i.e., the mean of the distribution is 36.6. The magnitude of the compromise effect can be influenced by application of an asymmetric value function after comparison of alternatives.

4. COMPARISON WITH OTHER MODELS

Multi-alternative DFT (Roe et al., 2001) and the LCA model (Usher and McClelland, 2004) both account for similarity, attraction, and compromise effects in three-alternative preferential choice and thus build the theoretical background for the $2N$-ary choice tree model. Nevertheless there are some important differences and the first one to set the new model apart from the previous approaches is the attribute-wise normalization of the
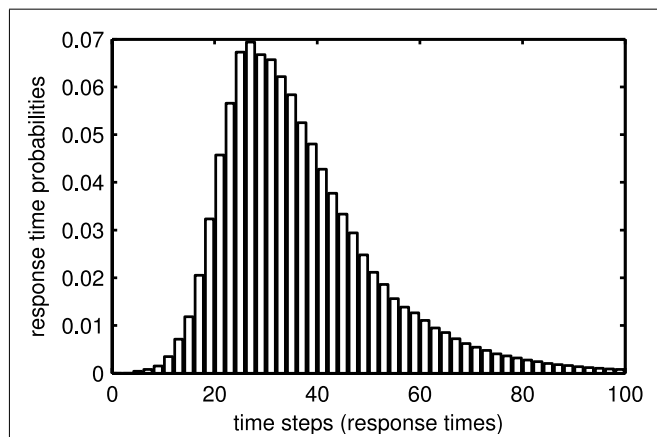

FIGURE 7 | Response time distribution for alternative $A = (9, 1)$ in the compromise setup with $\theta = 5$, $\xi = 0.01$, $\lambda = 0.05$, and $\mu = 0$. The expected response time, i.e., the mean of the distribution is 36.6.

initially provided evaluations of alternatives. This preprocessing of input values makes them comparable over attributes. Effects that originate from differing orders of magnitude of the input values can thus be controlled by influencing the attention weights for the attributes. The comparison of alternatives on single attributes is basically the same in all three models but only the LCA model and the $2N$-ary choice tree model allow for external reference points that are not present in the choice set to influence the resulting values. Application of an asymmetric value function allows the LCA model to implement the loss-aversion principle (Kahneman and Tversky, 1979) and addition of a positive constant avoids negative activations and thus negated inhibition which was crucial for some of the results of multi-alternative DFT. Both concepts (asymmetric value function and positive constant) can be implemented into the $2N$-ary choice tree model as well but do not affect its ability to account for the aforementioned effects (except for the magnitude of the compromise effect). Whereas all three models use leakage to account for decay of already sampled information over time and have a random part that implements noise in human decision making, inhibition is another crucial difference between them. In multi-alternative DFT, local inhibition explains the attraction and compromise effect, the LCA model uses global inhibition to account for the similarity effect. Both types of inhibition can be implemented in the $2N$-ary choice tree model but are not necessary for explanation of the three effects.

Beside some similarities and dissimilarities between the models, in particular with respect to some underlying psychological concepts the $2N$-ary choice tree model is the first to provide expected choice probabilities and response times in closed form and thus allows for convenient estimation of the model parameters from the observed choice times and frequencies in experimental settings. Furthermore, it can be extended to more than three-choice alternatives in a straightforward way to account for choice behavior in more complex, and possibly more realistic choice situations.

5. CONCLUDING REMARKS

The $2N$-ary choice tree model provides alternative explanations for the similarity, attraction, and compromise effect that can be experimentally tested as suggested before. Especially the manipulation of attention weights is of interest, because it differentiates the model on hand from former approaches and should allow to experimentally produce similarity and attraction effects which has been proven to be difficult in the past. One problem, however, we are currently encountering is limited machine accuracy which leads to accumulation of rounding errors during calculation of expected choice probabilities and response times.

6. FORMAL STATEMENT AND PROOF

We can approximate the expected choice probabilities and hitting times up to absolute accuracy in finite time. This follows from Theorem 1:

Theorem 1: Each random walk Y_n on the above defined tree $T = (V, E, r)$ with transition probabilities p_e, ends in finite time with probability one.

Corollary 1: With probability one only finitely many addends in equations (1) and (2) are unequal zero.

Considering expected values, i.e., limits of infinite sums, it is helpful to make use of a concept that allows for propositions about asymptotic behavior. For each alternative, the difference of the two counters that are associated with this alternative resemble a birth-death chain:

Definition 4 (birth-death chain): A sequence of random variables $X_1, X_2, \ldots$ with values in a countable state space $\mathscr{S} \equiv \{0, 1, 2, \ldots\} \subseteq \mathbb{N}$ is called *Markov chain*, if it satisfies the Markov property

$$\mathbb{P}\left[X_{n+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\right]$$
$$= \mathbb{P}\left[X_{n+1} = x | X_n = x_n\right].$$

A Markov chain is called *time homogeneous*, if $\mathbb{P}[X_{n+1} = x | X_n = y] = p(x, y)$ for all n, i.e., the probability for going from x to y is independent from n. A *birth-death chain* is a time homogeneous Markov chain that does not skip any state. Its transition probabilities $p(x, y)$ are equal to zero for all $x, y \in \mathscr{S}$ with $|y - x| > 1$. A *non-homogeneous birth-death chain* is a birth-death chain that is not time homogeneous.

Durrett (2010) proves the following theorem for (non-homogeneous) birth-death chains as special case of Markov chains:

Theorem 2: Let X_n be a Markov chain and suppose

$$\mathbb{P}\left[\cup_{m=n+1}^{\infty} \{X_m \in B_m\} | X_n\right] \geq \delta > 0 \text{ on } \{X_n \in A_n\}.$$

Then $\mathbb{P}[\{X_n \in A_n \ \textit{infinitely} \ \textit{often}\} - \{X_n \in B_n \ \textit{infinitely} \ \textit{often}\}] = 0$.

For each alternative $i \in \{1, 2, \ldots, N\}$, the above mentioned difference of its two counters can be interpreted as non-homogeneous birth-death chain X_n with absorbing states θ_i^- and θ_i^+ and state space $\mathscr{S} = \{\theta_i^-, (\theta_i^- + 1), (\theta_i^- + 2), \ldots, (\theta_i^+ - 2), (\theta_i^+ - 1), \theta_i^+\}$. Its transition probabilities are

$$\left.\begin{array}{ll} p^n(x, x+1) & = p_x^n = p_i^n(x), \\ p^n(x, x-1) & = q_x^n = p_{N+i}^n(x), \\ p^n(x, x) & = r_x^n = 1 - p_x^n - q_x^n, \end{array}\right\} \quad \text{for } x \in \mathscr{S} - \left\{\theta_i^-, \theta_i^+\right\}$$

and

$$\left.\begin{array}{ll} p_x^n = q_x^n & = 0, \\ r_x^n & = 1, \end{array}\right\} \quad \text{for } x \in \left\{\theta_i^-, \theta_i^+\right\}.$$

Due to the noise in the transition probabilities, $p_x^n > 0$ and $q_x^n > 0$ for all $x \in \mathscr{S} - \{\theta_i^-, \theta_i^+\}$. It follows that the probability for walking the direct way from $x \in \mathscr{S} - \{\theta_i^-, \theta_i^+\}$ to either θ_i^- or θ_i^+ is

$$\delta_x := \left(\prod_{y=\theta_i^- + 1}^{x} q_i^{n+x-y}(y) + \prod_{z=x}^{\theta_i^+ - 1} p_i^{n+z-x}(z)\right) > 0$$

and thus

$$\delta := \min_{\theta_i^- < x < \theta_i^+} \delta_x > 0.$$

Define $T_{\theta_i^-} = \inf\{n : X_n = \theta_i^-\}$, $T_{\theta_i^+} = \inf\{n : X_n = \theta_i^+\}$, $T_i = T_{\theta_i^-} \wedge T_{\theta_i^+}$, $A_n = \mathscr{S}$, and $B_n = \{\theta_i^-, \theta_i^+\}$. Then $\{X_m \in B_m\}$ is equivalent to $T_i \leq m$ and

$$\mathbb{P}\left[\cup_{m=n+1}^{\infty} \{X_m \in B_m\} | X_n\right] = 1 > \delta$$

for $X_n \in B_n$. The probability for walking from any $x \in \mathscr{S}$ to either θ_i^- or θ_i^+ on every possible way is

$$\mathbb{P}\left[\cup_{m=n+1}^{\infty} \{X_m \in B_m\} | X_n\right] \geq \delta > 0$$

and thus fulfills the assumptions of Theorem 2. It follows that

$$\mathbb{P}\left[\left\{X_n \in A_n \text{infinitely often}\right\} - \left\{X_n \in B_n \text{infinitely often}\right\}\right] = 0$$

which is equivalent to

$$\mathbb{P}\left[\left\{X_n \in A_n - B_n \text{ finitely often}\right\}\right] = \mathbb{P}\left[T_i < \infty\right] = 1$$

and as this is true for every alternative $i \in \{1, 2, \ldots, N\}, \mathbb{P}[T, \infty] = 1$ holds for $T := \min T_1, T_2, \ldots, T_N$. This proves Theorem 1.

REFERENCES

Busemeyer, J. R., and Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Math. Soc. Sci.* 23, 255–282.

Busemeyer, J. R., and Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychol. Rev.* 100, 432–455.

Diederich, A. (1995). Intersensory facilitation of reaction time: evaluation of counter and diffusion coactivation models. *J. Math. Psychol.* 39, 197–215.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *J. Math. Psychol.* 41, 260–274.

Diederich, A. (2008). A further test of sequential-sampling models that account for payoff effects on response bias in perceptual decision tasks. *Percept. Psychophys.* 70, 229–256.

Diederich, A., and Busemeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: bound-change, drift-rate-change, or two-stage-processing hypothesis. *Percept. Psychophys.* 68, 194–207.

Durrett, R. (2010). *Probability: Theory and Examples*, 4th Edn. New York: Cambridge University Press.

Hick, W. E. (1952). On the rate of gain of information. *Q. J. Exp. Psychol.* 4, 11–26.

Huber, J., Payne, J. W., and Puto, C. (1982). Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis. *J. Consum. Res.* 9, 90–98.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *J. Exp. Psychol.* 45, 188–196.

Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291.

Kahneman, D., and Tversky, A. (1991). Loss aversion in riskless choice: a reference-dependent model. *Q. J. Econ.* 106, 1039–1061.

Koehler, D. J., and Harvey, N. (2007). *Blackwell Handbook of Judgment and Decision Making*, 2nd Edn. Oxford: Blackwell Publishing.

Korte, B., and Vygen, J. (2002). *Combinatorial Optimization: Theory and Algorithms*, 2nd Edn. Berlin: Springer.

LaBerge, D. (1962). A recruitment theory of simple behavior. *Psychometrika* 27, 375–396.

LaBerge, D. (1994). Quantitative models of attention and response processes in shape identification tasks. *J. Math. Psychol.* 38, 198–243.

Laming, D. R. J. (1968). *Information Theory of Choice Reaction Times*. New York: Academic Press.

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

Luce, D. (1986). *Response Times*. New York: Oxford University Press.

Pike, A. R. (1966). Stochastic models of choice behaviour: response probabilities and latencies of finite Markov chain systems. *Br. J. Math. Stat. Psychol.* 19, 15–32.

Ratcliff, R., and Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychol. Rev.* 111, 333–367.

Roe, R. M., Busemeyer, J. R., and Townsend, J. T. (2001). Multialternative decision field theory: a dynamic connectionist model of decision making. *Psychol. Rev.* 108, 370–392.

Simonson, I. (1989). Choice based on reasons: the case of attraction and

compromise effects. *J. Consumer Res.* 16, 158–174.

Smith, P. L., and Van Zandt, T. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *Br. J. Math. Stat. Psychol.* 53, 293–315.

Townsend, J. T., and Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Psychological Processes*. London: Cambridge University Press.

Tversky, A. (1972). Choice by elimination. *J. Math. Psychol.* 9, 341–367.

Tversky, A., and Simonson, I. (1993). Context-dependent preferences. *Manage. Sci.* 39, 1179–1189.

Usher, M., and McClelland, J. L. (2001). On the time course of perceptual choice: the leaky competing

accumulator model. *Psychol. Rev.* 108, 550–592.

Usher, M., and McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychol. Rev.* 111, 757–769.

Van Zandt, T., Colonius, H., and Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychon. Bull. Rev.* 7, 208–256.

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.