Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

LOOKAHEAD CONTRACTION POLICIES FOR BAYESIAN RANKING AND SELECTION WITH PAIRWISE COMPARISONS

Laura Priekule
Stephan Meisel

School of Business and Economics
University of Muenster
Leonardo-Campus 3
48149 Muenster, GERMANY

ABSTRACT

We propose and evaluate novel sampling policies for a Bayesian ranking and selection problem with pairwise comparisons. We introduce the lookahead contraction principle and apply it to three types of value factors for lookahead policies. The resulting lookahead contraction policies are analyzed both with the minimal number of lookahead steps required for obtaining informative value factors, and with fixed number of lookahead steps. We show that lookahead contraction reduces the minimal number of required lookahead steps, and that contraction guarantees finiteness of the minimal lookahead. For minimal lookahead we demonstrate empirically that lookahead contraction never leads to worse performance, and that lookahead contraction policies based on expected value of improvement perform best. For fixed lookahead, we show that all lookahead contraction policies eventually outperform their counterparts without contraction, and that contraction results in a performance boost for policies based on predictive probability of improvement.

1 INTRODUCTION

In this work we propose and evaluate novel sampling policies for a Bayesian ranking and selection (R&S) problem with pairwise comparisons. In particular, we focus on lookahead policies for this problem, i.e., on policies that make sampling decisions based on the hypothetical assumption that a number of pairwise comparisons of any two alternatives can be observed in the next sampling stage. For such policies, we introduce the principle of lookahead contraction, which concentrates the predicted effect of several consecutive pairwise comparisons of one pair to a single pairwise comparison. In the considered R&S problem we have a finite set of alternatives, where the quality of each alternative is unknown initially, but can be estimated by collecting information through pairwise comparisons with any other alternative. We define an alternative's quality by its average probability of winning against all other alternatives in the considered set of alternatives. This quality measure is equivalent to the well-known *Borda score* (de Borda 1784), and does not impose any assumptions on the winning probabilities. Our goal is to determine the best alternative by efficiently allocating a limited sampling budget across the pairs in a sequential procedure.

Pairwise comparisons provide a straightforward assessment approach, if on the one hand it is difficult to elicit an absolute scale of the alternatives' qualities, and if on the other hand relative judgments or preferences ("A is better than B") are available. The research interest on utilization of information from pairwise comparisons was initially driven by applications in social sciences (David 1988) and sports (Elo 1978). More recently, online gaming (Herbrich et al. 2007), as well as applications regarding information retrieval, such as crowdsourcing, or web search and recommender systems (Radlinski et al. 2008) have become of interest. In addition, pairwise comparisons arise in emerging fields such as use of batteries on energy markets. In the literature, the main focus concerning sequential sampling with pairwise comparisons

978-1-7281-3283-9/19/$31.00 ©2019 IEEE

lies on finding a complete ranking or on finding the top *k* alternatives (see, e.g., Heckel et al. (2018)), as well as on the so called dueling bandit problem, which is an extension of the multi-armed bandit problem to the case of preference-based feedback (Busa-Fekete and Hüllermeier 2014). However, the problem of finding the best alternative via R&S with pairwise comparisons has received far less attention.

In our R&S problem a pairwise comparison between two alternatives generates information in terms of a sample from a Bernoulli distribution. Under this assumption, however, a single pairwise comparison does often not yield enough information to derive good sampling decisions. As a consequence, especially sampling policies with a one-step lookahead may become ineffective, i.e., the policies may frequently not be able to allocate a sample deliberately. Kamiński (2015) discusses this issue for the knowledge gradient (KG) policy (Gupta and Miescke 1996; Frazier et al. 2008) in the context of a standard R&S problem (without pairwise comparisons) with Bernoulli samples. To mitigate the issue, he proposes to use a multi-step lookahead policy and minimizes the lookahead steps needed to acquire a sufficient amount of information. The proposed policy is based on an approximation of the predictive probability of a change of the alternative that is considered best, and is shown to outperform KG with one-step lookahead.

Groves and Branke (2019) encounters the issue of too little information in the context of solving a top-*k* selection problem with pairwise comparisons. The authors propose a one-step lookahead policy that is based on the predictive probability of a change of the *k* alternatives that are considered best. Whenever the proposed policy is unable to make a sampling decision, it is replaced by a pure exploration policy. The policy is shown to outperform state-of-the-art policies for top-*k* selection problems.

Priekule and Meisel (2017) considers a R&S problem with pairwise comparisons that is closely related to our problem. In order to mitigate negative impact of the lack of information from a single observation, the authors propose a variant of the KG* policy. KG* (Frazier and Powell 2010) is a multi-step lookahead policy based on the average expected value of information per step. The proposed policy is shown to perform better than the one-step lookahead KG policy, but becomes computationally demanding even for a relatively small number of alternatives, and falls back to pure exploration repeatedly.

In this work, we propose the principle of lookahead contraction for R&S with pairwise comparisons. We apply lookahead contraction to three different classes of multistep lookahead policies, which we refer to as predictive probability of improvement, predictive probability of identity change, and expected value of improvement. We show analytically that lookahead policies with contraction are never worse than their counterparts without contraction in terms of probability of not being able to make a deliberate sampling decision. Additionally, we propose lookahead contraction policies where this probability equals zero, and we show empirically that lookahead contraction improves the ability of Bayesian lookahead policies to cope with the sparse information coming from binary pairwise comparisons.

The remainder of this paper is structured as follows: In Section 2 we introduce the sampling model of our R&S problem with pairwise comparisons, and we propose to use lookahead policies for making sampling decisions. In Section 3 we introduce the principle of lookahead contraction and derive value factors for three classes of lookahead policies, both with and without contraction. In Section 4 we first derive the minimal number of lookahead steps that our policies without contraction require to minimize the probability of being unable to calculate a deliberate sampling decision, and we then derive the minimal number of lookahead steps that our policies with contraction require to guarantee the ability to calculate deliberate sampling decisions. In Section 5 we compare the performances of the policies with and without lookahead contraction empirically. Section 6 concludes the paper.

2 PROBLEM FORMULATION

In Section 2.1, we define the considered sampling model for pairwise comparisons, and in Section 2.2 we introduce the notion of lookahead policies for making sampling decisions.

2.1 Sampling Model for Pairwise Comparisons

Suppose that there are M alternatives. We adopt the multibinomial model, i.e., for each alternative $i = 1,\dots,M$ we define a vector $p_i = (p_{i1}, p_{i2}, \dots, p_{iM})$ of unknown winning probabilities, such that for any pair (i, j) with $i, j = 1,\dots,M$ we may observe pairwise comparisons in terms of random variables $W_{ij}^1, W_{ij}^2, \dots$, whose realizations are drawn independently from a Bernoulli distribution with success probability p_{ij}. We require that each pairwise comparison results in a winner, and thus $p_{ij} + p_{ji} = 1$ is always satisfied. For the sake of notational simplicity and without loss of generality, we set $p_{ii} = 0.5$. We measure the quality of alternative i in terms of its average winning probability

$$\bar{p}_i = \frac{1}{M}\sum_{j=1}^{M} p_{ij},$$

which is equivalent to the Borda score (de Borda 1784). The best alternative i^* is defined as the alternative with the largest average winning probability, i.e., $i^* \in \arg\max_{i\in M} \bar{p}_i$.

We learn about the unknown winning probabilities (and thus about the average winning probabilities) by making a sequence of N pairwise comparisons, i.e., at each stage $n = 1,\dots,N-1$, we sample a pair $(i, j)^n$ and observe $W_{(i,j)^n}^{n+1}$, which equals 1 if i wins over j and 0 otherwise. Adopting a Bayesian framework, we use the information collected from the sampling process $(i, j)^0, W_{(i,j)^0}^1, (i, j)^1, W_{(i,j)^1}^2, \dots, (i, j)^{n-1}, W_{(i,j)^{n-1}}^n$ to form the stage n beliefs about our values of interest in terms of probability distributions. In particular, our beliefs about the winning probabilities p_{ij} at stage n are beta distributed, i.e., $p_{ij} \sim \text{Beta}(\alpha_{ij}^n, \beta_{ij}^n)$ and the symmetry yields $p_{ji} \sim \text{Beta}(\beta_{ij}^n, \alpha_{ij}^n)$. Without loss of generality, we let $\alpha_{ii}^n = \beta_{ii}^n = 1$ for all $i \in M$. At stage $n+1$, our beliefs are updated according to the Bayesian equations

$$\alpha_{ij}^{n+1} = \begin{cases} \alpha_{ij}^n + W_{ij}^{n+1} & \text{if } (i,j)^n = (i,j) \\ \alpha_{ij}^n & \text{otherwise,} \end{cases}$$

$$\beta_{ij}^{n+1} = \begin{cases} \beta_{ij}^n + (1 - W_{ij}^{n+1}) & \text{if } (i,j)^n = (i,j) \\ \beta_{ij}^n & \text{otherwise.} \end{cases}$$

Note that without loss of generality, we assume that the set of pairs considered for sampling is defined as $\widetilde{M} = \{(i, j) \mid 1 \le i < j \le M\}$, with $|\widetilde{M}| = (M^2 - M)/2$. As a consequence the updating equations apply to all pairs (i, j) satisfying $j > i$, and the beliefs regarding the remaining pairs are updated using the symmetry.

Throughout this work, we follow the convention that any quantity indexed by n becomes known or can be computed at stage n. Furthermore, we write $\mathbb{P}^n$ and $\mathbb{E}^n$ to indicate probabilities and expectations with respect to distributions induced by the beliefs at stage n. Let $p_{ij}^n = \mathbb{E}^n[p_{ij}]$ be the posterior mean of the winning probability of pair (i, j) after n pairwise comparisons. And let $n_{ij}^n = \alpha_{ij}^n + \beta_{ij}^n$, such that $p_{ij}^n = \alpha_{ij}^n / n_{ij}^n$. Then the expected average winning probability of an alternative i after n comparisons can be calculated with respect to the posterior distribution as

$$\bar{p}_i^n := \mathbb{E}^n[\bar{p}_i] = \frac{1}{M}\sum_{j=1}^{M} p_{ij}^n = \frac{1}{M}\sum_{j=1}^{M} \frac{\alpha_{ij}^n}{n_{ij}^n}.$$

We denote the estimated value of the best alternative according to our beliefs at stage n as $\bar{p}_*^n = \max_{i\in M} \bar{p}_i^n$. And, we suppose that after exhausting the budget of N samples, a risk-neutral decision maker selects alternative $i^N = \arg\max_{i\in M} \bar{p}_i^N$.

2.2 Lookahead Policies

Let $S^n := (\alpha^n, \beta^n)$ denote the beliefs at stage n, where $\alpha^n = (\alpha_{ij}^n)_{M\times M}$ and $\beta^n = (\beta_{ij}^n)_{M\times M}$. We assume that our sampling decisions are controlled by a policy X^π mapping the beliefs S^n to a pair $X^\pi(S^n) \in \widetilde{M}$.

Following a policy X^π implies that, at each sampling stage $n < N$, a value factor $v_{ij}^{\pi,n}$ is assigned to each considered pair (i,j), and that the sampling decision proposed by the policy is then given by

$$X^\pi(S^n) = \underset{(i,j)\in\widetilde{M}}{\arg\max}\, v_{ij}^{\pi,n}.$$

The explicit definition of value factors associates a policy a unique assessment of the potential of each pair (i,j) to improve the current estimate $\bar{p}_*^n$. The aim generally is to choose a policy that helps to identify the truly best alternative in an efficient manner.

Lookahead policies use value factors that are defined based on the hypothetical assumption that a number $\tau_n \geq 1$ of pairwise comparisons of the chosen pair (i,j) can be observed in the next stage $n+1$, where we refer to τ_n as the lookahead at stage n. Note that in case of pairwise comparisons an advantage of lookahead policies over policies such as probability of improvement (Kushner 1964) and expected improvement (Jones et al. 1998) is that the calculations of value factors are more easily manageable. In the following sections we consider lookahead policies with value factors that are defined in terms of either probabilities or expectations.

3 VALUE FACTORS FOR LOOKAHEAD POLICIES

In Section 3.1 we introduce the principle of lookahead contraction, and in Section 3.2 we define three different types of value factors for lookahead policies. In Section 3.3 we combine these two steps by proposing three different types of value factors with lookahead contraction.

3.1 Lookahead Contraction

We introduce lookahead contraction as a modification of the change we predict at stage n for the expected average winning probabilities. Using a lookahead policy we have at stage n a predictive distribution for the estimates $\bar{p}^{n+1}$ under the hypothetical assumption that we observe $\tau_n \geq 1$ pairwise comparisons of alternatives i and j in the next stage $n+1$. In particular, the posterior distribution at n induces a predictive distribution for W_{ij}^{n+1}, which is Beta-binomial, i.e.,

$$\mathbb{P}^n(W_{ij}^{n+1} = k) = \binom{\tau_n}{k} \frac{B(\alpha_{ij}^n + k, \beta_{ij}^n + \tau_n - k)}{B(\alpha_{ij}^n, \beta_{ij}^n)},$$

where $B(\cdot,\cdot)$ is the beta function. Furthermore, τ_n pairwise comparisons of i and j imply predicted changes of the expected average winning probabilities $\bar{p}_i$ and $\bar{p}_j$. These changes are given by

$$\hat{p}_{ij}(W_{ij}^{n+1}, \tau_n) = \frac{W_{ij}^{n+1} - p_{ij}^n \tau_n}{M(n_{ij}^n + \tau_n)}, \tag{1}$$

such that $\bar{p}_i^{n+1} = \bar{p}_i^n + \hat{p}_{ij}(W_{ij}^{n+1}, \tau_n)$ and $\bar{p}_j^{n+1} = \bar{p}_j^n - \hat{p}_{ij}(W_{ij}^{n+1}, \tau_n)$. The predicted change $\hat{p}_{ij}(W_{ij}^{n+1}, \tau_n)$ is a discrete random variable with zero expectation. In addition, the function $\hat{p}_{ij}(k, \tau_n)$ is monotonically increasing in k, and thus, for fixed τ_n, the change $\hat{p}_{ij}$ is bounded by $\hat{p}_{ij}^n(0, \tau_n)$ and by $\hat{p}_{ij}(\tau_n, \tau_n)$. These bounds define the range of observable values of the predicted change, which we denote as the *effect range* of our τ_n pairwise comparisons. The effect range is growing as τ_n increases, but is bounded by the limits $\lim_{\tau_n\to\infty} \hat{p}_{ij}(0, \tau_n) = -p_{ij}^n/M$ and $\lim_{\tau_n\to\infty} \hat{p}_{ij}(\tau_n, \tau_n) = p_{ji}^n/M$.

Note that lookahead policies make sampling decisions based on the two contradicting principles of (a) the assumption that several consecutive pairwise comparisons of a pair are made at one stage, and (b) the fact that only one single pairwise comparison is actually made at one stage. Against this background we propose to transform the predicted change given by Equation (1) into the modified predicted change

$$\hat{p}_{ij}^C(W_{ij}^{n+1}, \tau_n) = \frac{n_{ij}^n + \tau_n}{n_{ij}^n + 1} \hat{p}_{ij}(W_{ij}^{n+1}, \tau_n) = \frac{W_{ij}^{n+1} - p_{ij}^n \tau_n}{M(n_{ij}^n + 1)}. \tag{2}$$

The modified predictive change $\hat{p}_{ij}^C$ allows for the following interpretation: The numerator of the fraction in Equation (2) accounts for the assumption of making τ_n pairwise comparisons and complies with the numerator of the original $\hat{p}_{ij}$ for lookahead policies. The denominator, however, differs from the original by reckoning only the number n_{ij}^n of pairwise comparisons of i and j “made so far” plus the single pairwise comparison that will be made in the next sampling step. Hence, we may think of a contraction of the impacts of τ_n sequential comparisons into the impact of one comparison. We refer to lookahead policies with value factors using the modified predicted change as lookahead contraction policies.

3.2 Value Factors without Lookahead Contraction

The value of a possible future sampling decision is typically derived either from (a) the decision’s potential to increase the expected performance of the alternative that is considered best, i.e., the potential to lead to positive $\Delta_*^{n+1} = \bar{p}_*^{n+1} - \bar{p}_*^n$, or from (b) the decision’s potential to lead to a change of the identity of the alternative that is considered best (which obviously depends on Δ_*^{n+1}). We consider the following three types of factors for calculating the value of a sampling decision $(i,j)^n$ at stage n with lookahead τ_n:

- The *predictive probability of improvement* (PPI) represents the value factor for a pair (i,j) in terms of the probability of an increase in the value of the alternative that is considered best, i.e.,

$$v_{ij}^{PPI,n}(\tau_n) = \mathbb{P}^n\left(\Delta_*^{n+1} > 0 \,\middle|\, (i,j)^n = (i,j), \tau_n\right).$$

- The *predictive probability of identity change* (PIC) represents a value factor in terms of the probability that the identity of the alternative that is considered best will change, and is given by

$$v_{ij}^{PIC,n}(\tau_n) = \mathbb{P}^n\left(i_*^{n+1} \neq i_*^n \,\middle|\, (i,j)^n = (i,j), \tau_n\right).$$

- The *expected value of improvement* (EVI), represents a value factor in terms of the expected change in the value of the alternative that is considered best, and is given by

$$v_{ij}^{EVI,n}(\tau_n) = \mathbb{E}^n[\Delta_*^{n+1} | (i,j)^n = (i,j), \tau_n]. \tag{3}$$

  Note that in case of $\tau_n = 1$ for all stages n, the EVI-policy is equivalent to Knowledge Gradient policy proposed by Frazier et al. (2008), and by Gupta and Miescke (1996).

We discuss similarities and differences of PPI, PIC, and EVI by reformulating the value factors with the predicted change $\hat{p}_{ij}(W_{ij}^n, \tau_n)$, as given in Equation (1). First, we rewrite Δ_*^{n+1} as

$$\Delta_*^{n+1}(\hat{p}_{ij}(W_{ij}^{n+1}, \tau_n)) = \max\left\{C_{ij}^n, \bar{p}_i^n + \hat{p}_{ij}(W_{ij}^{n+1}, \tau_n), \bar{p}_j - \hat{p}_{ij}(W_{ij}^{n+1}, \tau_n)\right\} - \bar{p}_*^n, \tag{4}$$

where $C_{ij}^n = \max_{l \neq i,j} \bar{p}_l^n$. This allows us to derive the value factors of PPI and PIC as follows:

Lemma 1 Define the constants $\Delta_i^n = \bar{p}_*^n - \bar{p}_i^n$. Then, for each pair $(i,j) \in \widetilde{M}$

$$v_{ij}^{PPI,n}(\tau_n) = 1 - \mathbb{P}^n\left(-\Delta_j^n \leq \hat{p}_{ij}(W_{ij}^{n+1}, \tau_n) \leq \Delta_i^n\right).$$

Lemma 2 Let $\Delta_{ij}^n = \bar{p}_n^* - C_{ij}^n$. Then for each pair $(i,j) \in \widetilde{M}$

$$v_{ij}^{PIC,n}(\tau_n) = \begin{cases} \mathbb{P}^n\left(\hat{p}_{ij}(W_{ij}^{n+1}, \tau_n) < -\min\{\frac{\Delta_j^n}{2}, \Delta_{ij}^n\}\right) & \text{if } i_*^n = i, \\ \mathbb{P}^n\left(\hat{p}_{ij}(W_{ij}^{n+1}, \tau_n) > \min\{\frac{\Delta_i^n}{2}, \Delta_{ij}^n\}\right) & \text{if } i_*^n = j, \\ 1 - \mathbb{P}^n\left(-\Delta_j^n \leq \hat{p}_{ij}(W_{ij}^{n+1}, \tau_n) \leq \Delta_i^n\right) & \text{otherwise.} \end{cases}$$

Lemmas 1 and 2 show that the value factors of PPI and PIC are equal whenever neither of the two alternatives of the pair being assessed is currently identified as best. Note that in this case positive Δ_*^{n+1} can only be achieved if the predicted change exceeds one of the constants, $-\Delta_j^n$ and Δ_i^n, which we denote as *lower* and *upper impact bound*. In case of $i_*^n \in \{i, j\}$, PPI and PIC differ in terms of the impact bounds. The PIC factor involves only one single impact bound (lower or upper, depending on which of i or j is currently considered as the best), as only a sufficiently large negative change of the expected performance of the alternative that is considered best is of significance. Furthermore, PIC has the same impact bounds as EVI. We show this by using Equations (3) and (4) to write the EVI factors as

$$v_{ij}^{EVI,n}(\tau_n) = \mathbb{E}^n[\Delta_*^{n+1}(\hat{p}_{ij}(W_{ij}^{n+1}, \tau_n))],$$

and by reformulating the EVI factors as stated in the following Lemma 3.

Lemma 3 For each pair $(i, j) \in \widetilde{M}$

- If $i_*^n = i$, then

$$\begin{aligned} v_{ij}^{EVI,n}(\tau_n) =& \mathbb{E}^n\left[\mathbb{1}_{\{\hat{p}_{ij}(W_{ij}^{n+1},\tau_n)<-\min\{\Delta_j^n/2,\Delta_{ij}^n\},\hat{p}_{ij}(W_{ij}^{n+1},\tau_n)<\bar{p}_j^n-C_{ij}^n\}}(-2\hat{p}_{ij}(W_{ij}^{n+1}, \tau_n) - \Delta_j^n)\right] \\ &+ \mathbb{E}^n\left[\mathbb{1}_{\{\bar{p}_j^n-C_{ij}^n\le\hat{p}_{ij}(W_{ij}^{n+1},\tau_n)<-\min\{\Delta_j^n/2,\Delta_{ij}^n\}\}}(-\hat{p}_{ij}(W_{ij}^{n+1}, \tau_n) - \Delta_{ij}^n)\right]. \end{aligned}$$

- If $i_*^n = j$, then

$$\begin{aligned} v_{ij}^{EVI,n}(\tau_n) =& \mathbb{E}^n\left[\mathbb{1}_{\{\hat{p}_{ij}(W_{ij}^{n+1},\tau_n)>\min\{\Delta_i^n/2,\Delta_{ij}^n\},\hat{p}_{ij}(W_{ij}^{n+1},\tau_n)>\bar{p}_i^n-C_{ij}^n\}}(2\hat{p}_{ij}(W_{ij}^{n+1}, \tau_n) - \Delta_i^n)\right] \\ &+ \mathbb{E}^n\left[\mathbb{1}_{\{\bar{p}_i^n-C_{ij}^n\ge\hat{p}_{ij}(W_{ij}^{n+1},\tau_n)>\min\{\Delta_i^n/2,\Delta_{ij}^n\}\}}(\hat{p}_{ij}^n(W_{ij}^{n+1}, \tau_n) - \Delta_{ij}^n)\right]. \end{aligned}$$

- Otherwise,

$$\begin{aligned} v_{ij}^{EVI,n}(\tau_n) =& \mathbb{E}^n\left[\mathbb{1}_{\{\hat{p}_{ij}(W_{ij}^{n+1},\tau_n)<-\Delta_j^n\}}(-\hat{p}_{ij}(W_{ij}^{n+1}, \tau_n) - \Delta_j^n)\right] \\ &+ \mathbb{E}^n\left[\mathbb{1}_{\{\hat{p}_{ij}(W_{ij}^{n+1},\tau_n)>\Delta_i^n\}}(\hat{p}_{ij}(W_{ij}^{n+1}, \tau_n) - \Delta_i^n)\right]. \end{aligned}$$

Note that Lemma 3 also shows that $0 \le v_{ij}^{EVI,n}(\tau_n) \le v_{ij}^{PIC,n}(\tau_n)$, where equality is the case if and only if $v_{ij}^{PIC,n}(\tau_n) = 0$.

3.3 Value Factors with Lookahead Contraction

Given the formulations of value factors in Lemmas 1–3, it is now straightforward to derive value factors with lookahead contraction. We substitute the predicted change with the modified predicted change as given in Equation (2), and define the cPPI, cPIC, and cEVI value factors with lookahead contraction as

$$\begin{aligned} v_{ij}^{cPPI(\tau_n),n} &= 1 - \mathbb{P}^n\big(-\Delta_j^n \le \hat{p}_{ij}^C(W_{ij}^{n+1}, \tau_n) \le \Delta_i^n\big), \\ v_{ij}^{cPIC(\tau_n),n} &= \begin{cases} \mathbb{P}^n\big(\hat{p}_{ij}^C(W_{ij}^{n+1}, \tau_n) < -\min\{\frac{\Delta_j^n}{2}, \Delta_{ij}^n\}\big) & \text{if } i_*^n = i, \\ \mathbb{P}^n\big(\hat{p}_{ij}^C(W_{ij}^{n+1}, \tau_n) > \min\{\frac{\Delta_i^n}{2}, \Delta_{ij}^n\}\big) & \text{if } i_*^n = j, \\ 1 - \mathbb{P}^n\big(-\Delta_j^n \le \hat{p}_{ij}^C(W_{ij}^{n+1}, \tau_n) \le \Delta_i^n\big) & \text{otherwise.} \end{cases}, \\ v_{ij}^{cEVI(\tau_n),n} &= \mathbb{E}^n[\Delta_*^{n+1}(\hat{p}_{ij}^C(W_{ij}^{n+1}, \tau_n))]. \end{aligned}$$

The value factors with lookahead contraction differ from the corresponding value factors without contraction in terms of their capacity to gain information from the (hypothetical) lookahead observations.

In particular, it can be shown easily that, for fixed τ_n and (i,j), the value factor with contraction is always greater than or equal to the corresponding value factor without contraction. Note that this results directly from the fact that $|\hat{p}^C_{ij}(k,\tau_n)| \geq |\hat{p}_{ij}(k,\tau_n)|$, where equality is the case if and only if $k = p^n_{ij}\tau_n$. Hence, the lookahead contraction essentially leads to an expansion of the effect range (cf. Section 3.1) of the pairwise comparisons. Moreover, the effect range expands as τ_n increases (as in the case without lookahead contraction) and is unbounded (in contrast to the case without lookahead contraction).

4 INFORMATIVE VALUE FACTORS FOR LOOKAHEAD POLICIES

The binary nature of the observations in our sampling process with pairwise comparisons may lead to situations where too little information is revealed in a stage of the process, i.e., where the value factors of our sampling policy equal zero. We refer to a value factor that equals zero as *uninformative*, and to a value factor that is greater than zero as *informative*. If in a stage n all value factors of our policy are uninformative, the policy looses its ability to deliberately allocate sampling decisions. In such a situation a common approach is to temporarily switch to a purely random exploration policy for the current stage n.

For each of the value factors introduced in Section 3 the chance of being informative depends directly on whether or not the factor's impact bounds lie within the effect range. Since the effect range expands as the lookahead τ_n increases, there is, for each value factor, a minimal lookahead required such that the factor becomes informative. We derive the minimal lookahead for value factors without lookahead contraction in Section 4.1, and for value factors with contraction in Section 4.2.

4.1 Minimal Lookahead without Contraction

For $\pi \in \{PPI, PIC, EVI\}$ the *minimal lookahead* for pair (i,j) is defined as

$$\tau_n^{ij,\pi} = \min\{\tau \in \mathbb{N} | v_{ij}^{\pi,n}(\tau) > 0\},$$

which implies that for each $\tau \geq \tau_n^{ij,\pi}$ the corresponding value factor $v_{ij}^{\pi,n}(\tau)$ is informative. We can derive the minimal lookahead for PPI and PIC as given by the following two lemmas.

Lemma 4 For each pair $(i,j) \in \widetilde{M}$, we have

$$\tau_n^{ij,PPI} = \begin{cases} 1 & \text{if } \Delta^n_j = 0 \text{ or } \Delta^n_i = 0, \\ \lceil \delta^-_{ij} \rceil & \text{if } \bar{p}^n_* = C^n_{ij},\ \Delta^n_j < \frac{p^n_{ij}}{M} \text{ and } \Delta^n_i \geq \frac{p^n_{ji}}{M}, \\ \lceil \delta^+_{ij} \rceil & \text{if } \bar{p}^n_* = C^n_{ij},\ \Delta^n_j \geq \frac{p^n_{ij}}{M} \text{ and } \Delta^n_i < \frac{p^n_{ji}}{M}, \\ \min\{\lceil \delta^-_{ij} \rceil, \lceil \delta^+_{ij} \rceil\} & \text{if } \bar{p}^n_* = C^n_{ij} \text{ and } \Delta^n_j < \frac{p^n_{ij}}{M} \text{ or } \Delta^n_i < \frac{p^n_{ji}}{M}, \\ \infty & \text{otherwise,} \end{cases}$$

where $\delta^-_{ij} = n^n_{ij}[(p^n_{ij}/M)(\Delta^n_j)^{-1} - 1]^{-1}$, and $\delta^+_{ij} = n^n_{ij}[(p^n_{ji}/M)(\Delta^n_i)^{-1} - 1]^{-1}$.

Lemma 5 For each pair $(i,j) \in \widetilde{M}$, we have

$$\tau_n^{ij,PIC} = \begin{cases} \tau_n^{ij,PPI} & \text{if } i^n_* \neq i, j, \\ 1 & \text{if } i^n_* = i \text{ and } \min\{\Delta^n_j, \Delta^n_{ij}\} = 0, \text{ or } i^n_* = j \text{ and } \min\{\Delta^n_i, \Delta^n_{ij}\} = 0, \\ \lceil \delta_i \rceil & \text{if } i^n_* = i \text{ and } \frac{p^n_{ij}}{M} > \min\{\Delta^n_j/2, \Delta^n_{ij}\} > 0, \\ \lceil \delta_j \rceil & \text{if } i^n_* = j \text{ and } \frac{p^n_{ji}}{M} > \min\{\Delta^n_i/2, \Delta^n_{ij}\} > 0, \\ \infty & \text{otherwise,} \end{cases}$$

where $\delta_i = n^n_{ij}[(p^n_{ij}/M)(\min\{\Delta^n_j/2, \Delta^n_{ij}\})^{-1} - 1]^{-1}$, and $\delta_j = n^n_{ij}[(p^n_{ji}/M)(\min\{\Delta^n_i/2, \Delta^n_{ij}\})^{-1} - 1]^{-1}$.

Lemmas 4 and 5 show that the minimal lookaheads of PPI and PIC are equal, if none of the two alternatives i and j is currently considered best, or if $\arg\max\{\bar{p}^n_i, \bar{p}^n_j, C^n_{ij}\}$ contains more than one index. In all other cases, the minimal lookahead of PPI always equals one, whereas the minimal lookahead of PIC tends to grow with increasing n^n_{ij} and may even get infinite.

The fact that the PIC value factors and the corresponding EVI value factors have the same impact bounds implies that the minimal lookahead of EVI is the same as the minimal lookahead of PIC, i.e.,

$$\forall(i,j) : \tau^{ij,EVI}_n = \tau^{ij,PIC}_n.$$

Hence, Lemmas 4 and 5 reveal that all three types of value factors may require an infinite lookahead in order to become informative. With PIC and EVI this may even happen for all pairs. Note that the possible occurrence of an infinite minimal lookahead results from the fact, that the effect range of the predictive change $\hat{p}_{ij}$ is bounded by the limiting effect range $(-p^n_{ij}/M, p^n_{ji}/M)$. An infinite minimal lookahead occurs whenever the impact bounds lie outside of the limiting effect range. We conclude with the general observation that the main driving force behind very large minimal lookaheads is given in terms of small ratios between the upper (lower) bound of the limiting effect range and the upper (lower) impact bound.

4.2 Minimal Lookahead with Contraction

Lookahead contraction implies that the impact bounds of the value factors for PPI, PIC and EVI (as given in Lemmas 1, 2, and 3) are multiplied with $(n^n_{ij}+1)/(n^n_{ij}+\tau_n)$. One consequence of this impact bound reduction is that the minimal lookahead becomes smaller if it was larger than 1 without lookahead contraction, and that it remains 1 otherwise. As in the case without lookahead contraction, the minimal lookaheads of EVI and of PIC with contraction are the same, i.e.,

$$\forall(i,j) : \tau^{ij,cEVI}_n = \tau^{ij,cPIC}_n.$$

We can derive the minimal lookaheads for cPPI and cPIC as given by the following two lemmas.

Lemma 6 For each pair $(i,j) \in \widetilde{M}$, we have

$$\tau^{ij,cPPI}_n = \begin{cases} 1 & \text{if } \Delta_i = 0 \text{ or } \Delta_j = 0, \\ \min\left\{\operatorname{ceil}\left(\frac{(n^n_{ij}+1)\Delta^n_j}{p^n_{ij}/M}\right), \operatorname{ceil}\left(\frac{(n^n_{ij}+1)\Delta^n_i}{p^n_{ji}/M}\right)\right\} & \text{otherwise,} \end{cases}$$

Lemma 7 For each pair $(i,j) \in \widetilde{M}$, we have

$$\tau^{ij,cPIC}_n = \begin{cases} \operatorname{ceil}\left(\frac{(n^n_{ij}+1)\min\{\Delta^n_j/2, \Delta^n_{ij}\}}{p^n_{ij}/M}\right) & \text{if } i^n_* = i, \\ \operatorname{ceil}\left(\frac{(n^n_{ij}+1)\min\{\Delta^n_j/2, \Delta^n_{ij}\}}{p^n_{ji}/M}\right) & \text{if } i^n_* = j, \\ \tau^{ij,cPPI}_n & \text{otherwise.} \end{cases}$$

By comparing Lemmas 4 and 5 with the corresponding Lemmas 6 and 7, we observe that lookahead contraction guarantees finiteness of the minimal lookahead. This means that lookahead contraction policies with minimal lookahead are always able to deliberately select a pair. In addition, it can be shown that lookahead contraction reduces the probability of uninformative value factors for all $\tau_n > 1$.

5 COMPUTATIONAL EXPERIMENTS

In this section we demonstrate by computational experiments the impact of lookahead contraction on the performance of lookahead policies. We describe our experimental setup in Section 5.1, and show numerical results in Section 5.2.

5.1 Experimental Setup

We report on two main experiments. First we consider problems with 10, 15 and 20 alternatives to illustrate the impact of lookahead contraction on the performances of policies with minimal lookahead. Second we use the case with 20 alternatives to demonstrate the impact of lookahead contraction on the performances of policies with fixed lookahead. For the first main experiment we define the lookahead policies PPI(min), PIC(min) and EVI(min), and their lookahead contraction counterparts cPPI(min), cPIC(min) and cEVI(min) as follows. For $\pi \in \{PPI, PIC, EVI, cPPI, cPIC, cEVI\}$ we let

$$X^{\pi(min)}(S^n) = \underset{(i,j)\in\widetilde{M}}{\arg\max} \frac{v_{ij}^{\pi,n}(\tau_n^{ij,\pi})}{\tau_n^{ij,\pi}}. \tag{5}$$

For the second experiment, we define policies with fixed lookaheads $\tau \in \{10, 40, 160\}$ for all n. We again consider $\pi \in \{PPI, PIC, EVI, cPPI, cPIC, cEVI\}$ and let

$$X^{\pi(\tau)}(S^n) = \underset{(i,j)\in\widetilde{M}}{\arg\max}\, v_{ij}^{\pi,n}(\tau). \tag{6}$$

If at a stage n all value factors of a policy equal zero, we temporarily substitute the policy with a purely random exploration policy. For each $M \in \{10, 15, 20\}$ we consider both a set of problems with weakly informative priors, and a set of problems with strongly informative priors. For each problem set we generate 10 initial priors by randomly choosing α_{ij}^0 and β_{ij}^0 (except for α_{ii}^0 and β_{ii}^0) from the sets $A_5 = \{k|k \leq 5, k \in \mathbb{N}\}$ and $A_{10} = \{k|k \leq 10, k \in \mathbb{N}\}$. Subsequently, we randomly sample 10 different truths with $p_{ij} \sim B(\alpha_{ij}^0, \beta_{ij}^0)$ per initial prior beliefs. Hence, each of the two considered problem sets consists of 10^2 randomly generated problem instances. In all our experiments we work with a sampling budget of $N = 1000$ pairwise comparisons. We measure the performance of a policy at stage n in terms of the true rank of the alternative that is considered best according to the beliefs at n. We adopt the convention that the alternative with the highest true average winning probability has rank one.

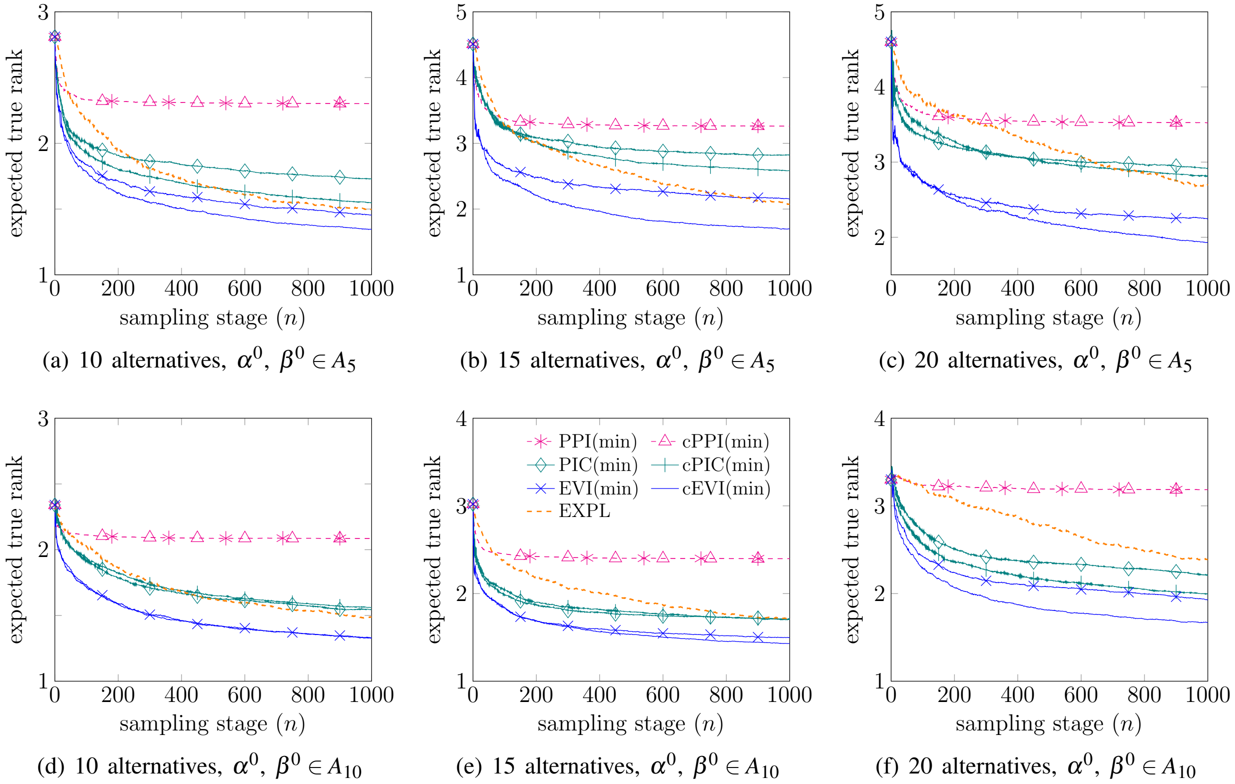
5.2 Results

Figure 1 displays the results of our first main experiment. For each of the policies defined in Equation (5) the figure shows the expected true rank of the alternative that is considered best according to the beliefs at stage n of the sampling process. The results of the pure exploration policy ("EXPL") are shown as benchmark. Figures 1a–1c show results for weakly informative initial priors $(\alpha^0, \beta^0 \in A_5)$, and Figures 1d–1f show results for strongly informative priors $(\alpha^0, \beta^0 \in A_{10})$. In all cases the expected true rank is calculated as average over $10^2 \times 100$ replications of the sampling process.

Figure 1 shows that in terms of the finally reached expected true rank none of the policies with lookahead contraction performs worse than its counterpart without contraction. In particular both cPIC(min) and cEVI(min) lead in most cases to significant performance gains compared with their counterparts without lookahead contraction, and, most prominently, we observe that cEVI(min) outperforms all other policies in each case. Figure 1 also reveals that the performance gains of cPIC(min) and cEVI(min) with respect to the counterparts PIC(min) and EVI(min) is larger in the presence of weakly informative initial priors.

We observe that the performances of PPI(min) and cPPI(min) are identical in each of the considered cases, and that the performance of these policies is always significantly worse than the performances of the other policies. We suppose that the fact that the performances are identical is a result of both policies choosing pairs with a minimal lookahead of 1, i.e., pairs containing the alternative that is currently considered best, which is the case where the value factors with and without contraction are equal.

Finally, Figure 1 shows that for minimal lookahead the three considered value factor types always lead to the same order of both the policies with contraction and the policies without contraction, i.e., in terms



(a) 10 alternatives, α^0, $\beta^0 \in A_5$

(b) 15 alternatives, α^0, $\beta^0 \in A_5$

(c) 20 alternatives, α^0, $\beta^0 \in A_5$

(d) 10 alternatives, α^0, $\beta^0 \in A_{10}$

(e) 15 alternatives, α^0, $\beta^0 \in A_{10}$

(f) 20 alternatives, α^0, $\beta^0 \in A_{10}$

Figure 1: Expected true rank of the alternative considered best according to the beliefs at stage n, for experiments with 10, 15 and 20 alternatives, and with weakly and strongly informative initial priors.

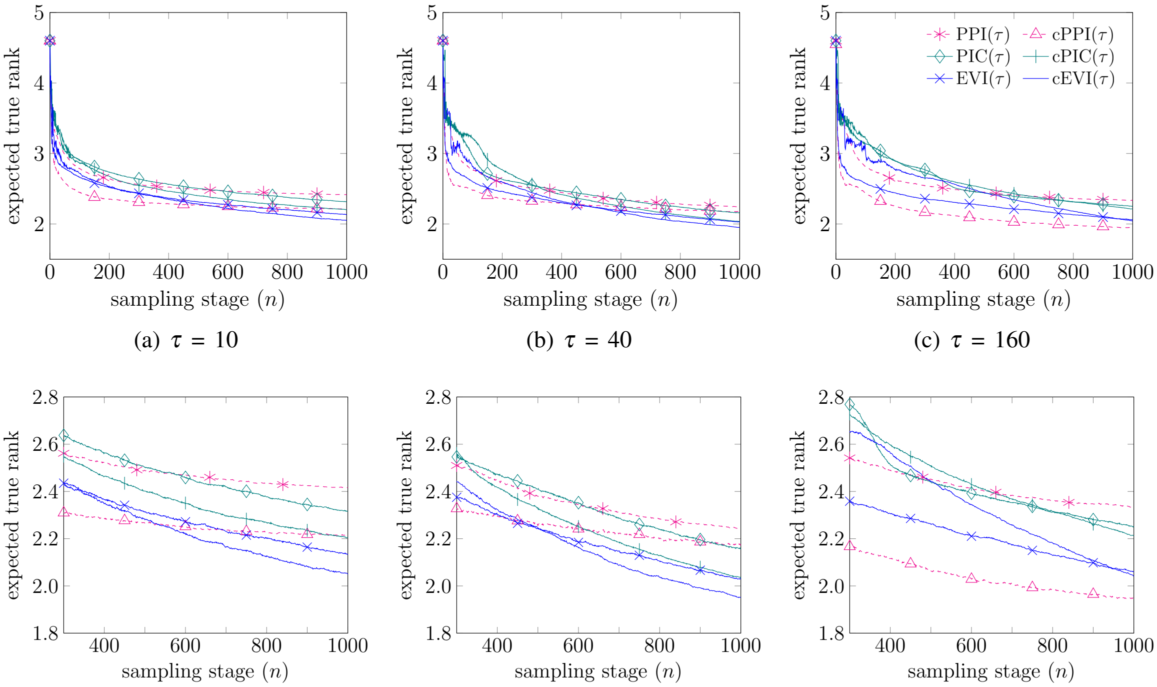
of expected true rank we consistently observe EVI $\prec$ PIC $\prec$ PPI, and cEVI $\prec$ cPIC $\prec$ cPPI. Our second experiment reveals that these orders may change if value factors are calculated with a fixed lookahead.

In our second main experiment, we focus on the case of $M = 20$ alternatives with weakly informative priors. For each of the policies defined in Equation (6), Figure 2 displays the expected true rank of the alternative that is considered best according to the beliefs at stage n of the sampling process. Figures 2a–2c show results for $\tau \in \{10, 40, 160\}$, and Figures 2d–2f represent magnifications of these results. In all cases the expected true rank is calculated as average over $10^2 \times 10^3$ replications of the sampling process.

The most prominent difference between the results of Figure 2, and the corresponding results with minimal lookahead (shown in Figure 1c) is that fixing the lookahead improves both PPI and cPPI significantly, and that the improvement of cPPI is larger than the improvement of its counterpart without lookahead contraction. The magnifications in Figures 2d–2f highlight that cPIC(τ) and cEVI(τ) outperform their counterparts without contraction for $\tau = 10$ and $\tau = 40$. Moreover, for large lookahead ($\tau = 160$), (i) the advantage of cPPI(τ) over PPI(τ) is particularly large, and (ii) cPIC(τ) and cEVI(τ) improve over PIC(τ) and EVI(τ) only towards the end of the sampling process. However, the slopes of cPIC(τ) and cEVI(τ) seem to indicate that the two policies have a significant advantage over PIC(τ) and EVI(τ) on the long run.

Note that in contrast to all lookahead contraction policies with minimal lookahead, both cPIC(τ), and cEVI(τ) are not guaranteed to avoid temporarily falling back to pure random exploration. Figure 3 illustrates the impact of the lookahead contraction on the according exploration rates and highlights that lookahead contraction reduces the exploration rate in all of the cases.

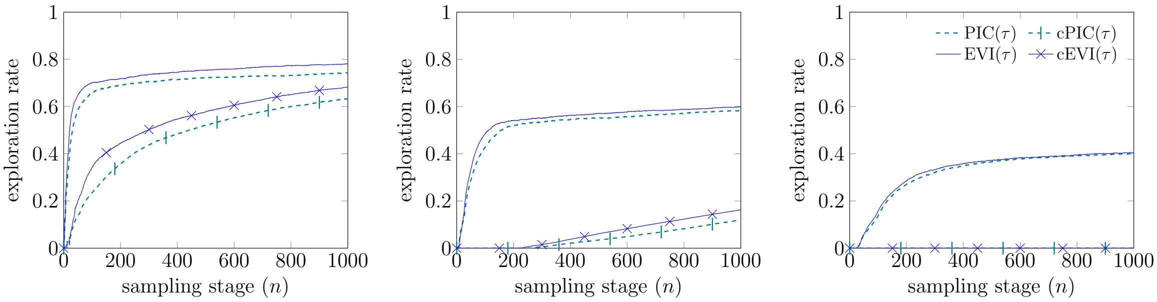
Preliminary experiments seem to indicate that, as the number of alternatives grows, the cPIC(τ) policy shows an increasing advantage also over other policies, such as the computationally intense KG* policy.



(a) τ = 10 (b) τ = 40 (c) τ = 160

(d) Magnification of Figure (2a) (e) Magnification of Figure (2b) (f) Magnification of Figure (2c)

Figure 2: Expected true rank of the best believed alternative, with 20 alternatives, α^0, $\beta^0 \in A_5$ and fixed τ



(a) τ = 10 (b) τ = 40 (c) τ = 160

Figure 3: Exploration rates at each stage of the sampling process, with 20 alternatives and α^0, $\beta^0 \in A_5$.

6 CONCLUSIONS

We introduce the lookahead contraction principle and propose a new class of lookahead policies for Bayesian ranking and selection with pairwise comparisons. Lookahead contraction improves the ability of lookahead policies to utilize the sparse information coming from binary observations by contracting the effects of several consecutive pairwise comparisons into one single pairwise comparison. We apply lookahead contraction to three different types of value factors for sampling policies: predictive probability of improvement, predictive probability of identity change, and expected value of improvement. We show that lookahead contraction reduces the minimum number of lookahead steps required to attain informative value factors, and that contraction guarantees finiteness of the minimal lookahead.

Our computational results demonstrate the advantage of lookahead contraction. For the case of policies using the minimal lookahead required for informative value factors, we show empirically that applying lookahead contraction never leads to worse performance, and that lookahead contraction policies based on expected value of improvement are best. For the case of fixed lookahead, our empirical results show that lookahead contraction results in a significant performance boost in particular for policies based on predictive probability of improvement. Avenues for future work are the study of lookahead contraction policies within different application contexts, and development of formal proofs of the advantage of lookahead contraction.

ACKNOWLEDGMENTS

The work of the authors was supported by the German Federal Ministry for Economic Affairs and Energy under grant 03ET6152 within the 6th Energy Research Programme of the Federal Government of Germany.

REFERENCES

Busa-Fekete, R., and E. Hüllermeier. 2014. "A Survey of Preference-Based Online Learning with Bandit Algorithms". In *International Conference on Algorithmic Learning Theory*, edited by P.Auer, A.Clark, T. Zeugmann, and S. Zilles, 18–39. Cham: Springer.

David, H. A. 1988. *The Method of Paired Comparisons*. 2nd ed. London: Hodder Arnold.

de Borda, J. C. 1784. "Mémoire sur les Élections au Scrutin". *Histoire de l'Academie Royale des Sciences pour 1781*:657–665.

Elo, A. E. 1978. *The Rating of Chessplayers, Past and Present*. New York: Arco Publishing Company.

Frazier, P. I., and W. B. Powell. 2010. "Paradoxes in Learning and the Marginal Value of Information". *Decision Analysis* 7(4):378–403.

Frazier, P. I., W. B. Powell, and S. Dayanik. 2008. "A Knowledge-Gradient Policy for Sequential Information Collection". *SIAM Journal on Control and Optimization* 47(5):2410–2439.

Groves, M., and J. Branke. 2019. "Top-κ Selection with Pairwise Comparisons". *European Journal of Operational Research* 274(2):615–626.

Gupta, S., and K. Miescke. 1996. "Bayesian Look Ahead One-Stage Sampling Allocations for Selection of the Best Population". *Journal of Statistical Planning and Inference* 54(2):229–244.

Heckel, R., M. Simchowitz, K. Ramchandran, and M. J. Wainwright. 2018. "Approximate Ranking from Pairwise Comparisons". *arXiv preprint arXiv:1801.01253*.

Herbrich, R., T. Minka, and T. Graepel. 2007. "TrueSkill$^{\text{TM}}$: a Bayesian Skill Rating System". In *Advances in Neural Information Processing Systems*, edited by J. Platt, D. Koller, Y. Singer, and S. Roweis, 569–576. San Diego: Neural Information Processing Systems Foundation, Inc.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-box Functions". *Journal of Global Optimization* 13(4):455–492.

Kamiński, B. 2015. "Refined Knowledge-Gradient Policy for Learning Probabilities". *Operations Research Letters* 43(2):143–147.

Kushner, H. J. 1964. "A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise". *Journal of Basic Engineering* 86(1):97–106.

Priekule, L., and S. Meisel. 2017. "A Bayesian Ranking and Selection Problem with Pairwise Comparisons". In *Proceedings of the 2017 Winter Simulation Conference*, edited by V. Chan, A. D'Ambrogio, G. Zacharewicz, and N. Mustafee, 2149–2160. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Radlinski, F., M. Kurup, and T. Joachims. 2008. "How Does Clickthrough Data Reflect Retrieval Quality?". In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, edited by J. Shanahan, S. Amer-Yahia, Y. Zhang, A. Kokz, A. Chowdury, and D. Kelly, 43–52. New York: Association for Computing Machinery.

AUTHOR BIOGRAPHIES

LAURA PRIEKULE is a Ph.D. student in the School of Business and Economics at the University of Muenster, Germany. She holds a German Diploma in Mathematics from University of Dresden, Germany. Her research interests lie in sequential decision-making under uncertainty with applications in energy. Her email address is laura.priekule@uni-muenster.de.

STEPHAN MEISEL is an Assistant Professor in the School of Business and Economics at the University of Muenster, Germany. He holds a Ph.D. in Operations Research, and was a postdoc at Princeton University. His research interests lie in sequential decision-making under uncertainty with applications. His email address is stephan.meisel@uni-muenster.de.