# Cumulative Dominance and Heuristic Performance in Binary Multiattribute Choice

## Manel Baucells
Department of Managerial Decision Sciences, IESE Business School, 08034 Barcelona, Spain,
mbaucells@iese.edu

## Juan A. Carrasco
Departament d'Enginyeria Electrònica, Universitat Politècnica de Catalunya, 08028 Barcelona, Spain,
carrasco@eel.upc.edu

## Robin M. Hogarth
Department of Economics and Business, ICREA and Universitat Pompeu Fabra, 08005 Barcelona, Spain,
robin.hogarth@upf.edu

We study the effectiveness of simple heuristics in multiattribute decision making. We consider the case of an additive separable utility function with nonnegative, nonincreasing attribute weights. In this case, cumulative dominance ensures that the so-called cumulative dominance compliant heuristics will choose a best alternative. For the case of binary attribute values and under two probabilistic models of the decision environment generalizing a simple Bernoulli model, we obtain the probabilities of simple and cumulative dominance. In contrast with the probability of simple dominance, the probability of cumulative dominance is shown to be large in many cases, explaining the effectiveness of cumulative dominance compliant heuristics in those cases. Additionally, for the subclass of the so-called fully cumulative dominance compliant heuristics, we obtain an upper bound for the expected loss that only depends on the weights being nonnegative and nonincreasing. The low values of the upper bound for cases in which the probability of cumulative dominance is not large provide an additional explanation for the effectiveness of fully cumulative dominance compliant heuristics. Examples of cumulative dominance compliant heuristics and fully cumulative dominance compliant heuristics are discussed, including the deterministic elimination by aspects (*DEBA*) heuristic that motivated our work.

## 1. Introduction

We consider an additive-separable multiattribute choice problem having $m$ alternatives $i$, $1 \leqslant i \leqslant m$, each characterized by $k$ attribute values $x_{i,r}$, $1 \leqslant r \leqslant k$. The utility of the $i$th alternative, with profile $x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,k})$, is defined as

$$U_i = w_1 x_{i,1} + w_2 x_{i,2} + \cdots + w_k x_{i,k}, \tag{1}$$

where the $w_r$ are nonnegative weighting parameters subject to the constraint $w_1 + w_2 + \cdots + w_k = 1$. As is customary, we assume that the $x_{i,r}$s have been normalized, and take values in the interval $[0, 1]$. The problem is to identify which of the $m$ alternatives is best, i.e., has the largest value of $U_i$. We assume here that the utilities $U_i$ are merely *ordinal*, that is, these values serve only to rank the alternatives $i$, and utility differences or averages have no necessary meaning as they would for cardinal or vNM utility. Equivalently, the utilities $U_i$ can be outputs of an additively separable value function (Keeney and Raiffa 1976) whose

marginal value functions yield the profile values $x_{i,j}$. The assumptions needed for this decomposition are appropriate or a good approximation in many situations.

In applications, it is often the case that the weighting parameters are not known with any precision. This therefore raises the issue of how sensitive decisions are to the lack of such knowledge. In this paper, we consider situations where although precise values are unknown, the decision maker can nonetheless order the weights by size. Thus, without loss of generality, we assume that $w_1 \geqslant w_2 \geqslant \cdots \geqslant w_k \geqslant 0$, and that the decision maker knows this correct ordering. This means that beginning with $x_1 = x_2 = \cdots = x_k = 0$, it is most preferred to swing $x_1$ to one, next most preferred to swing $x_2$ to one, and so on. This assumption is realistic in many scenarios. In our teaching experience, managers strongly agree with the claim that prioritizing is an essential managerial skill that many people perform regularly in their decision making. There is a considerable body of literature on using additive value functions to rank order alternatives using these same assumptions (Barron

and Barrett 1996). For example, Butler et al. (1997) discuss this problem and use sensitivity analysis for the weights, and Rao and Sobel (1980) use the rank order to derive marginal distributions for the weights.

The range of applications is broad because it includes instances of both individual and group decision making. Consider having to choose one of several candidates to fill a job opening, or one of several locations for a vacation. It is easy to imagine a decision maker, or a group, that may easily know or agree on which attributes are the most relevant, and on their order. However, this individual or group may have difficulties in determining or reaching agreement on, the precise values to be given to the weights. Of course, members must also agree on the precise values of each alternative on each attribute.

Rather than attempting to find precise values of the weights or to reach a group agreement (Baucells and Sarin 2003), the individual or group might consider using a decision rule that does not depend on using exact values for the weights. In this paper, we explore how the use of a class of such decision rules or heuristics affects performance and thus contributes to gaining understanding when the knowledge of precise weights is or is not necessary for effective decision making. Specifically, we exploit the concept of *cumulative dominance* (Kirkwood and Sarin 1985), the core concept of this paper. To illustrate, consider two alternatives with attribute profiles given by $x_1 = (0.8, 0.2, 0.5)$ and $x_2 = (0, 1, 0.4)$. Here, alternative 1 cumulative dominates alternative 2 because $x_{1,1} \geqslant x_{2,1}$, $x_{1,1} + x_{1,2} \geqslant x_{2,1} + x_{2,2}$ and $x_{1,1} + x_{1,2} + x_{1,3} \geqslant x_{2,1} + x_{2,2} + x_{2,3}$. Cumulative dominance is present when there is at least one alternative that cumulative dominates all others. If weights are nonnegative and nonincreasing, then any such alternative has the highest utility value. Of course, in the presence of simple dominance (an alternative is best on all attributes), then any such alternative will also have the highest utility.

Given a set of alternatives, one can easily check for the presence of simple or cumulative dominance. However, to study this issue with greater generality, one needs to specify the decision environment. To do so, we consider the attribute values of the $m$ alternatives as being the realization of a probabilistic model. We restrict attention to binary attributes, taking values zero and one. Although this is not the most general case of multiattribute choice, the binary setting does have an interest on its own. For example, it is common to have features that are either present or absent (e.g., the candidate has a certain academic degree or not), or that take two values (e.g., the candidate is male or female). Also, to simplify the decision, the decision maker may use a cutoff to classify the attribute values as either high ($=1$) or low ($=0$) (e.g., the candidate has a sufficient knowledge of a certain foreign language or not). Finally, even if the attribute is multivalued, the decision maker may be able to distinguish between zero and nonzero values, but be insensitive to the actual magnitude of the attribute (Hsee and Rottenstreich 2004).

We propose two probabilistic models that are generalizations of a simple Bernoulli model in which attribute values $x_{i,r}$ are assumed to be independent Bernoulli random variables with parameter $p$ (taking value one with probability $p$ and taking value zero with probability $1 - p$), $0 < p < 1$. In the first model, the probability $p$ may change from attribute to attribute. In the second model, we allow positive interattribute correlation. One of our goals is to calculate exact probability values for the presence of simple and cumulative dominance in these environments. For simple dominance, we derive a closed-form expression in the Bernoulli case, and recursion-based expressions for the other cases. For cumulative dominance, we apply ROBDDs, a well-known computational technique suited for binary settings, to calculate the probabilities. We are able to apply the ROBDD-based approach for up to 10 alternatives and 10 attributes, which is well beyond the scope of previous studies (Kirkwood and Corner 1993).

To facilitate comparison, we set as a base case the Bernoulli model with parameter $p = 1/2$. We then perform sensitivity analysis with respect to (i) high and low values of $p$; (ii) low-to-high and high-to-low patterns for $p_r$, $1 \leqslant r \leqslant k$; and (iii) moderate and high levels of correlation. As expected, the presence of simple dominance decays rapidly as the number of attributes grows. In contrast, the presence of cumulative dominance decays more slowly. Probabilities of cumulative dominance are high in favorable environments (few attributes, high $p$, and high correlation). Moreover, probabilities of cumulative dominance are reasonable in unfavorable environments (many attributes, low $p$, and low correlation). Being able to calculate exact probabilities for up to 10 alternatives provides a surprising result: as the number of alternatives increases, levels of simple and cumulative dominance fall at first, but then increase. We provide an explanation for this.

The high probabilities of cumulative dominance explain the effectiveness of certain heuristics whose key property is *compliance* with cumulative dominance: in the presence of cumulative dominance, the heuristic chooses (one of) the dominant alternatives. Hence, the probability of having cumulative dominance is a lower bound for the probability that a compliant heuristic chooses a best alternative.

A heuristic possessing this property, and the one that originally motivated this study, is deterministic elimination by aspects (*DEBA*), which is a variant of the elimination by aspects heuristic proposed by Tversky (1972). (In our formulation, the ordering in which attributes are considered is fixed, whereas in Tversky's model the ordering of attributes is random, with probabilities proportional to the weights. Our model, however, uses randomization to resolve possible ties among equivalent alternatives.) We therefore call it *DEBA*. *DEBA* is a lexicographic selection algorithm based on the attribute ordering. Specifically, at the first stage, unless all alternatives have the same values (all zeros or all ones in the binary case), alternatives with the nonhighest value (zero in the binary case) on the first attribute are

eliminated. If all have the same attribute values, then all alternatives remain at this stage. After this elimination, if a single alternative remains, it is chosen. Otherwise, the values of the second attribute are examined, and alternatives with the nonhighest values on that second attribute are eliminated. Note that *DEBA* cannot eliminate all alternatives because if all alternatives under consideration have the same value on a given attribute, then all alternatives survive this stage; otherwise, one or more alternatives necessarily have the highest attribute value and survive this stage. This procedure continues until only one alternative remains or all attributes have been examined. If several alternatives remain after all attributes have been examined, then their attribute values and utility are necessarily the same, and one of them is chosen at random. *DEBA* is easy to use because, in many situations, there is no need to look beyond the first or second attribute to make a decision.

Another important class of heuristics that is compliant with cumulative dominance is *EWn/DEBA*. *EWn/DEBA* first chooses the alternatives with the highest total sum of attribute values up to attribute *n*, and then breaks ties using *DEBA*.

The expected loss of a heuristic is defined as the expected difference between the utility of a best alternative and the utility of the alternative chosen by the heuristic. We provide an upper bound on the expected loss for those heuristics that are fully compliant with cumulative dominance, i.e., choose one of the alternatives in the set of alternatives that cumulative dominates all other alternatives up to the highest possible attribute index. *DEBA* is the only heuristic that we know in the literature that is fully compliant. We show that the upper bound for the expected loss remains reasonable in cases in which the probability of cumulative dominance is not high, reinforcing the justification for the observed effectiveness of *DEBA*. We emphasize that this upper bound for the expected loss is independent of the weights.

This paper has two main contributions. First, it demonstrates the key role that cumulative dominance has in the general study of heuristics. Second, based on cumulative dominance, it develops methodologies to calculate properties of certain classes of heuristics in binary environments. Taken together, we are able to justify the observed good performance of lexicographic *DEBA*-like heuristics (Hogarth and Karelaia 2005b, 2006a). As a by-product, we also develop methodologies to compute the probability of simple dominance in binary environments, establishing a lower bound for the probability of choosing a best alternative for most, if not all, heuristics.

The rest of this paper is organized as follows. In §2, we formally introduce simple and cumulative dominance, and discuss their role in multicriteria decision making. In this section, we also obtain an upper bound independent of the precise values of the weights for the loss (looked at as a random variable) for any fully cumulative dominant heuristic. Section 3 introduces the probabilistic models, explains first how to calculate the probabilities of having simple and cumulative dominance, second how to calculate the upper bound for the expected loss, and third reports specific numerical results. Section 4 discusses the performance of *DEBA* and other heuristics in relation to cumulative dominance. Section 5 concludes and highlights directions for future work. Most of the proofs are short and are kept in the main text. An appendix, available in the electronic companion to this paper, contains the proofs underlying the closed-form expression for the probability of simple dominance for the Bernoulli model and the recursion-based expressions for the probability of simple dominance for the other two probabilistic models. The electronic companion to this paper is available as part of the online version that can be found at http://or.pubs.informs.org/.

## 2. Compliance with Dominance

The definitions and results of this section do not depend on the attributes being binary. Unless stated otherwise, we assume that the instances of $x_{i,r}$ take real values.

### 2.1. Compliance with Simple Dominance

Alternative $i$ exhibits *simple dominance* over alternative $j$ up to attribute $r$, denoted by $s_r(i, j)$, if and only if $x_{i,t} \geqslant x_{j,t}$ for all $t$, $1 \leqslant t \leqslant r$. It is trivial to check that simple dominance characterizes optimality for nonnegative weights.

**Theorem 1.** $U_i \geqslant U_j$ *for all weights* $w_r \geqslant 0$, $1 \leqslant r \leqslant k$, *if and only if* $s_k(i, j)$.

We introduce the set $S_r$, which contains the alternatives that exhibit simple dominance over any other alternative up to attribute $r$. Formally,

$$S_r = \{1 \leqslant i \leqslant m \colon s_r(i, j), \ 1 \leqslant j \leqslant m\}.$$

Obviously, $S_1 \supset S_2 \supset \cdots \supset S_k$. Note that $S_1$ is nonempty, and all alternatives in $S_r$ have identical profiles up to attribute $r$. For example, in the case of three alternatives with profiles $x_1 = (1, 0, 1, 0)$, $x_2 = (1, 0, 0, 1)$, and $x_3 = (0, 1, 1, 0)$, $S_1 = \{1, 2\}$ and $S_2 = S_3 = S_4 = \varnothing$. More importantly, that $S_k$ is nonempty is an indicator of the presence of simple dominance, and the alternative(s) in $S_k$ will have the largest utility. A heuristic *complies with simple dominance* if it chooses an alternative from $S_k$ whenever $S_k$ is nonempty. Most heuristics, if not all, comply with simple dominance (this is trivial and we omit the proof of the following result).

**Theorem 2.** *For all weights* $w_r \geqslant 0$, $1 \leqslant r \leqslant k$, *if* $S_k$ *is nonempty, then any heuristic that complies with simple dominance will choose a best alternative.*

## 2.2. Compliance with Cumulative Dominance

To present the notion of cumulative dominance, we first consider the cumulative profile of an alternative $i$, $X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,k})$, where component $r$ is given by

$$X_{i,r} = \sum_{t=1}^{r} x_{i,t}.$$

Cumulative dominance is identical to simple dominance, but applied to the cumulative profile: Alternative $i$ exhibits *cumulative dominance* over alternative $j$ up to attribute $r$, denoted by $c_r(i,j)$, if and only if $X_{i,t} \geqslant X_{j,t}$ for all $t$, $1 \leqslant t \leqslant r$. Alternative $i$ cumulative dominates alternative $j$ if $i$ cumulative dominates $j$ up to attribute $k$, i.e., if $c_k(i,j)$. Figure 1 illustrates cumulative dominance using the same example as in the previous subsection. In the example of the figure, alternatives 1 and 2 cumulative dominate alternative 3 up to attribute 2, and alternative 1 cumulative dominates alternatives 2 and 3. Presenting Equation (1) as

$$U_i = (w_1 - w_2)X_{i,1} + (w_2 - w_3)X_{i,2}$$
$$+ \cdots + (w_{k-1} - w_k)X_{i,k-1} + w_k X_{i,k}, \quad (2)$$

we clearly see that because the weights are nonnegative and nonincreasing, a cumulative dominant alternative has the highest utility. In fact, cumulative dominance characterizes optimality for nonnegative, nonincreasing weights (Kirkwood and Sarin 1985).

THEOREM 3. $U_i \geqslant U_j$ for all weights $w_1 \geqslant w_2 \geqslant \cdots \geqslant w_k \geqslant 0$ if and only if $c_k(i,j)$.

PROOF. From Equation (2), it follows that

$$U_i - U_j = \sum_{t=1}^{k-1} (w_t - w_{t+1})(X_{i,t} - X_{j,t}) + w_k(X_{i,k} - X_{j,k}),$$

which is necessarily nonnegative if alternative $i$ cumulative dominates alternative $j$ and weights are nonnegative and nonincreasing. For the converse, recall that $\sum_{t=1}^{k} w_t x_{i,t} \geqslant \sum_{t=1}^{k} w_t x_{j,t}$ holds for all nonincreasing weights. Hence, it holds for $(w_1, w_2, w_3, \ldots, w_k)$ set to $(1, 0, 0, \ldots, 0)$, $(1/2, 1/2, 0, \ldots, 0), \ldots$, and $(1/k, 1/k, 1/k, \ldots, 1/k)$, respectively. These $k$ inequalities establish $c_k(i,j)$. $\square$

In parallel with simple dominance, we let $C_r$ denote the set of alternatives that exhibit cumulative dominance over any other alternative up to attribute $r$. Formally,

$$C_r = \{1 \leqslant i \leqslant m: c_r(i,j), 1 \leqslant j \leqslant m\}.$$

Obviously, $C_1 \supset C_2 \supset \cdots \supset C_k$. Note that $C_1$ is always nonempty, and that all alternatives in $C_r$ have identical profiles up to attribute $r$. In the example of Figure 1, $C_1 = C_2 = \{1, 2\}$ and $C_3 = C_4 = \{1\}$. Most importantly, if $C_k$ is nonempty, then Theorem 3 guarantees that the alternative(s) in $C_k$ will have the largest utility. Hence, choosing from $C_k$ is a desirable property. A decision heuristic is said to *comply with cumulative dominance* if it chooses an alternative from $C_k$ whenever $C_k$ is nonempty. This shows the following.

**Figure 1.** Cumulative profiles of three alternatives $x_1$, $x_2$, and $x_3$.
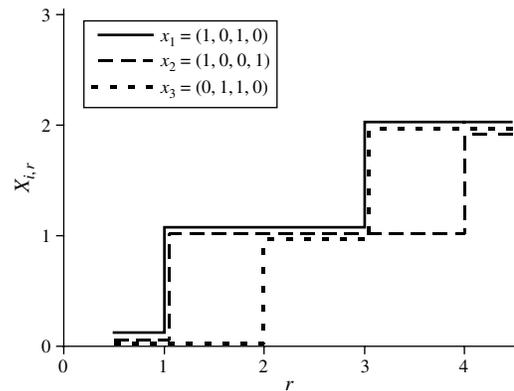


THEOREM 4. *For all weights* $w_1 \geqslant w_2 \geqslant \cdots \geqslant w_k \geqslant 0$, *if* $C_k$ *is nonempty, then any heuristic that complies with cumulative dominance will choose a best alternative.*

## 2.3. Full Compliance with Cumulative Dominance

Cumulative dominance may be lost at some attribute $r^*$, i.e., $C_{r^*} \neq \varnothing$ and $C_{r^*+1}, \ldots, C_k = \varnothing$. Formally, $r^*$ is the largest attribute index for which one alternative cumulative dominates all others:

$$r^* = \max_{1 \leqslant r \leqslant k} \{r: C_r \neq \varnothing\}.$$

The case $r^* = 1$ can only happen if there exist two alternatives $i, j$ with $x_{i,1} > x_{j,1}$ and $x_{i,1} + x_{i,2} < x_{j,1} + x_{j,2}$ (e.g., $x_1 = (0.5, 0, 0)$ and $x_2 = (0, 0.8, 0)$). If attributes are binary, then these two inequalities cannot be satisfied simultaneously, and $r^* \geqslant 2$.

A heuristic *fully complies with cumulative dominance* if it always chooses an alternative from $C_{r^*}$. Of course, full compliance with cumulative dominance implies compliance with cumulative dominance and is a desirable property because it ensures a small loss whenever $r^*$ is high. To see this, let $c$ be the alternative chosen by a certain heuristic. Given this choice, we define the loss as the difference between the utility of the best alternative and the utility of $c$, i.e.,

$$L = \max_{1 \leqslant i \leqslant m} U_i - U_c. \quad (3)$$

Note that value differences such as $L$ are not necessarily meaningful for ordinal utilities $U_i$. To make $L$ meaningful, it suffices to assume that the $U_i$ arise from an additively separable measurable value function (Dyer and Sarin 1979).

How large can $L$ be? Recall that a fully compliant heuristic will choose an alternative $c$ in $C_{r^*}$. Let $j$ be the alternative with highest utility. Compared to $j$, how large can the loss of $c$ be? Intuitively, the loss cannot be larger than when: (i) the cumulative profiles of $j$ and $c$ attain the same value at $r^*$ ($X_{j,r^*} = X_{c,r^*}$), (ii) $j$ is optimum past

$r^*$ $(x_{j,r^*+1} = x_{j,r^*+2} = \cdots = x_{j,k} = 1)$, (iii) $c$ is unfavorable past $r^*$ $(x_{c,r^*+1} = x_{c,r^*+2} = \cdots = x_{c,k} = 0)$, and (iv) the weights are all equal $(w_r = 1/k, 1 \leqslant r \leqslant k)$. In such a case, we obtain $U_j - U_c = (k - r^*)/k$. We can rigorously prove that $(k - r^*)/k$ is an upper bound for $L$ independent of the weights.

THEOREM 5. *If attributes take values between zero and one and weights are nonnegative and nonincreasing, then any heuristic that fully complies with cumulative dominance will have a loss with respect to the best alternative $\leqslant (k - r^*)/k$.*

PROOF. Given a set of weights and using (2), the loss of the alternative $c$ chosen by the heuristic with respect to any other alternative $i$, $U_i - U_c$, cannot be larger than in the case in which $i$ would meet the following two conditions: (1) $X_{i,t} = X_{c,t}$, $1 \leqslant t \leqslant r^*$ (because $c \in C_{r^*}$, $X_{i,t} \leqslant X_{c,t}$, $1 \leqslant t \leqslant r^*$), (2) $X_{i,t} = X_{c,t} + (t - r^*)$, $r^* + 1 \leqslant t \leqslant k$ (which is possible because all $x_{i,t}$, $r^* + 1 \leqslant t \leqslant k$ could be one and all $x_{c,t}$, $r^* + 1 \leqslant t \leqslant k$ could be zero). Again using (2), we have

$$L \leqslant \sum_{t=r^*+1}^{k-1} (w_t - w_{t+1})(t - r^*) + w_k(k - r^*) = \sum_{t=r^*+1}^{k} w_t,$$

which is zero for $r^* = k$. Assume that $1 \leqslant r^* < k$. To complete the proof, it suffices to show $\sum_{t=r^*+1}^{k} w_t \leqslant (k - r^*)/k$ in that case. The proof is by contradiction. Assume that $\sum_{t=r^*+1}^{k} w_t > (k - r^*)/k$. Because the weights are nonincreasing, we have $\sum_{t=r^*+1}^{k} w_t \leqslant (k - r^*)w_{r^*+1}$ and $w_{r^*+1} \geqslant (\sum_{t=r^*+1}^{k} w_t)/(k - r^*) > 1/k$. Again using the fact that the weights are nonincreasing, $\sum_{t=1}^{r^*} w_t \geqslant r^* w_{r^*+1} > r^*/k$ and, combining with $\sum_{t=r^*+1}^{k} w_t > (k - r^*)/k$, $\sum_{t=1}^{k} w_t > r^*/k + (k - r^*)/k = 1$, the sought contradiction. □

## 3. Probability of Dominance

In this section, we restrict ourselves to binary attribute values, i.e., $x_{i,r}$ is either zero or one, $1 \leqslant i \leqslant m$, $1 \leqslant r \leqslant k$.

### 3.1. How Pervasive Is Dominance?

For given $m$ and $k$, one can list all $2^{mk}$ possible combinations of profiles for the $m$ alternatives. Then, count the combinations for which one alternative dominates all others, either using simple dominance or cumulative dominance. This count divided by the $2^{mk}$ combinations agrees with the probability of simple and cumulative dominance, respectively, when all $x_{i,r}$ are independent Bernoulli random variables with $p = 0.5$. Table 1 reports these calculations for the case of $m = 2$ alternatives. In the table and henceforth, we write $P_S$ and $P_C$ to denote $P[S_k \neq \varnothing]$ and $P[C_k \neq \varnothing]$, respectively. In the case of $k = 3$, for example, cumulative dominance holds in 62 out of 64 possible ordered pairs. The exceptions are $x_1 = (1, 0, 0)$, $x_2 = (0, 1, 1)$, and vice versa.

Simple dominance is surprisingly common up to three attributes, but decays rapidly as $k$ increases. In contrast,

**Table 1.** Possible number of profiles of two alternatives, numbers of cases of simple and cumulative dominance, and corresponding probabilities $P_S$ or $P_C$, respectively.

| $k$ | $2^{2k}$ | # $S_k \neq \varnothing$ | # $C_k \neq \varnothing$ | $P_S$ | $P_C$ |
|---|---|---|---|---|---|
| 1 | 4 | 4 | 4 | 1.000 | 1.000 |
| 2 | 16 | 14 | 16 | 0.875 | 1.000 |
| 3 | 64 | 46 | 62 | 0.719 | 0.969 |
| 4 | 256 | 146 | 236 | 0.570 | 0.922 |
| 5 | 1,024 | 454 | 892 | 0.443 | 0.871 |
| 6 | 4,096 | 1,394 | 3,368 | 0.340 | 0.822 |
| 7 | 16,384 | 4,246 | 12,742 | 0.259 | 0.778 |
| 8 | 65,536 | 12,866 | 48,364 | 0.196 | 0.738 |

cumulative dominance is guaranteed for up to two attributes, and decays slowly with $k$. Even with eight attributes, we have cumulative dominance in more than 73% of the cases! We later show that $P_C$ in Table 1 goes to zero as $k$ increases. To have a better sense of the likely presence of dominance, we should explore cases with more than two alternatives. Better yet, we would like to use more sophisticated probabilistic models for the $x_{i,r}$. This is the goal of the probabilistic models that we present next.

### 3.2. Two Binary Probabilistic Models of the Environment

Decision makers live in an environment that determines the set of alternatives from which they choose. We use probabilistic models to capture such an environment: For fixed values of $m$ and $k$, the $k$ attribute values of the $m$ alternatives are realizations of some probability distribution. We propose two models that are generalizations of a simple Bernoulli model in which the $x_{i,r}$ are independent Bernoulli random variables with parameter $p$, $0 < p < 1$. Our first model is the ZIAC (Zero InterAttribute Correlation) model.

DEFINITION 1 (ZIAC MODEL). Given values of $p_r$, $0 < p_r < 1$, $1 \leqslant r \leqslant k$, the $x_{i,r}$ are independent Bernoulli random variables with parameter $p_r$ (taking value one with probability $p_r$ and value zero with probability $1 - p_r$).

The ZIAC model is a trivial generalization of the Bernoulli model. In the Bernoulli model, $E[x_{i,r}] = p$ and $p$ measures the average "quality" of any attribute. In the ZIAC model, $E[x_{i,r}] = p_r$ and $p_r$ measures the average quality of attribute $r$. Our second model, the PIAC (Positive InterAttribute Correlation) model, is also a generalization of the Bernoulli model.

DEFINITION 2 (PIAC MODEL). Given values of $p$, $1 < p < 1$, and $\rho$, $0 \leqslant \rho < 1$, the $x_{i,r}$ are obtained as $x_{i,r} = z y^h_{i,r} + (1 - z) y^l_{i,r}$, where the $z$, $y^h_{i,r}$, and $y^l_{i,r}$ are independent Bernoulli random variables with parameters $p$, $p^h = p + \sqrt{\rho}(1 - p)$, and $p^l = p - \sqrt{\rho}p$, respectively.

The PIAC model is intuitively appealing: what produces the positive correlation between any attribute value pair of

any alternative is a common cause that shifts the average quality of the attributes of the alternative. This is captured by the alternatives belonging to a "good" population with probability $p$ and to a "bad" population with probability $1 - p$. Alternatives in the good population follow a Bernoulli model with parameter $p^h = p + \sqrt{\rho}(1 - p)$, and alternatives in the bad population follow a Bernoulli model with parameter $p^l = p - \sqrt{\rho}p$. In the PIAC model, $E[x_{i,r}] = p$, so $p$ measures the average quality of any attribute and one can check that the positive correlation between any two attribute values of any given alternative is precisely $\rho$.

Before explicitly calculating $P_S$ and $P_C$ in these environments, there are certain limiting properties when either the number of alternatives or the number of attributes grows large. For fixed $k$, the probability, $P_1$, that some alternative will have all its attributes equal to one tends to one as $m \to \infty$, ensuring that, for fixed $k$, $P_S \to 1$ as $m \to \infty$. Of course, if $P_S$ tends to one, so does $P_C$.

**THEOREM 6.** *For both the ZIAC and the PIAC models, for fixed $k$, both $P_S$ and $P_C$ tend to one as $m$ tends to infinity.*

PROOF. That $P_1$ tends to one can be easily seen for the ZIAC model by noting that for that model, $P_1 = 1 - (1 - \prod_{r=1}^{k} p_r)^m$. For the PIAC model, it suffices to note that for that model,

$$P_1 = \sum_{g=0}^{m} \binom{m}{g} p^g (1-p)^{m-g} [1 - (1 - (p^h)^k)^g (1 - (p^l)^k)^{m-g}]$$

$$\geqslant \sum_{g=0}^{m} \binom{m}{g} p^g (1-p)^{m-g} [1 - (1 - (p^l)^k)^m]$$

$$= 1 - (1 - (p^l)^k)^m. \quad \square$$

Next, we consider the case of fixed $m$, and a number of attributes that increase. Because this case is more complex, in the case of $P_C$ we restrict ourselves to the Bernoulli model. Furthermore, for $P_S$ we impose some conditions on the ZIAC model.

**THEOREM 7.** *For the ZIAC model with a sequence of parameters $p_1, p_2, p_3, \ldots$ satisfying $\inf_{r \geqslant 1} p_r > 0$ and $\sup_{r \geqslant 1} p_r < 1$ and for the PIAC model, for fixed $m$, $P_S$ tends to zero as $k$ tends to infinity. For the Bernoulli model, for fixed $m$, $P_C$ tends to zero as $k$ tends to infinity.*

PROOF. That for both models, for fixed $m$, $P_S$ tends to zero as $k \to \infty$ follows from the fact that in both models, assuming $S_k$ nonempty, the probability that $|S_k|$ decreases when $k$ increases by one is bounded from below by a positive quantity $q$ independent of $k$. Hence, as $k$ increases, $|S_k|$ will decrease in a finite number of $k$ increments with probability one and, because $m$ is fixed, as $k$ increases, $S_k$ will become empty with probability one. To determine a $q > 0$ for the ZIAC model, note that the probability that $|S_k|$ will not decrease is the probability that either all alternatives have attribute $k + 1$ value zero or all alternatives in $S_k$

have attribute $k + 1$ value one, which is $(1 - p_{k+1})^m + p_{k+1}^{|S_k|} \leqslant (1 - p_{k+1})^m + p_{k+1}$. Hence, the probability that $|S_k|$ will decrease is $\geqslant 1 - p_{k+1} - (1 - p_{k+1})^m = (1 - p_{k+1}) \times [1 - (1 - p_{k+1})^{m-1}] \geqslant (1 - \sup_{r \geqslant 1} p_r)[1 - (1 - \inf_{r \geqslant 1} p_r)^{m-1}] > 0$. To determine a $q > 0$ for the PIAC model, note that $|S_k|$ decreases if $x_{i,k+1} = 0$ for, say, the minimum $i \in S_k$ and $x_{j,k+1} = 1$ for, say, the minimum $j \neq i$. Alternatives $i$ and $j$ will be good/good, good/bad, bad/good, and bad/bad with some probability mass function. Hence, the probability with which $|S_k|$ will decrease will be $\geqslant \min\{(1 - p^h)p^h, (1 - p^h)p^l, (1 - p^l)p^h, (1 - p^l)p^l\} = (1 - p^h)p^l > 0$.

To see that for the Bernoulli model, for fixed $m$, $P_C$ tends to zero as $k \to \infty$, note that

$$\{C_k \neq \varnothing\} = \bigcup_{n=1}^{m} \bigcap_{\substack{n'=1 \\ n' \neq n}}^{m} \{c_k(n, n')\}.$$

Hence, it suffices to prove $\lim_{k \to \infty} P[c_k(n, n')] = 0$, $n' \neq n$. But, for the Bernoulli model, $X_{n,k} - X_{n',k}$, $n' \neq n$ because a function of $k$ is a symmetric random walk. It is well known that a symmetric random walk crosses zero with probability one (Durret 2004, Chapter 3). Such crossing will break any cumulative dominance between alternatives $n$ and $n'$, implying $\lim_{k \to \infty} P[c_k(n, n')] = 0$, $n' \neq n$. $\square$

We calculate the probability of encountering simple and cumulative dominance in these environments. The computational procedures we develop apply to any desired parameter values. For illustrative purposes, we set as the base-case scenario the Bernoulli model with parameter $p = 0.5$. We then consider the following six variations.

(1) To examine the effect of the average attribute quality, we change the parameter of the Bernoulli model to the values $p = 0.2$ and $p = 0.8$.

(2) To examine the effect of varying increasing average attribute quality, we consider the ZIAC model with a *low-to-high pattern* in which $p_r$ increases linearly with $r$ from $p_1 = 0.2$ to $p_k = 0.8$, i.e., $p_r = 0.2 + 0.6(r - 1)/(k - 1)$. To examine the effect of varying decreasing average attribute quality, we consider a *high-to-low pattern* in which $p_r$ decreases linearly with $r$ from $p_1 = 0.8$ to $p_k = 0.2$, i.e., $p_r = 0.8 - 0.6(r - 1)/(k - 1)$. The average of the $p_r$s is in both cases 0.5, which allows us to compare the results with the base case.

(3) Finally, to examine the effect of positive interattribute correlation, we consider the PIAC model with $p = 0.5$ and two levels of correlation, $\rho = 0.2$ and $\rho = 0.5$.

### 3.3. Probability of Having Simple Dominance

We are able to find a closed-form expression for $P_S$ for the Bernoulli model. Throughout the paper, we use $x^0 = 1$, $x$ real $\neq 0$ (see, for example, Beardon 1979, pp. 55–56).

**THEOREM 8.** *For the Bernoulli model, the probability that one alternative exhibits simple dominance over all others is*

$$P_S = \sum_{i=1}^{m} \binom{m}{i} (-1)^{i-1} [p^i + (1 - p)^m]^k.$$

The appendix in the online supplement contains the proof. In this same appendix, we also obtain, using discrete-parameter Markov chains, recursion-based expressions to calculate $P_S$ for the ZIAC and PIAC models. Using these recursion-based expressions, we can calculate $P_S$ for quite large values of $k$ and $m$.

Figure 2 plots $P_S$ for values of $k$ ranging from 2 to 10 and values of $m$ ranging from 2 to 10. The seven graphs correspond to the base-case scenario and the six variations. The more important conclusion is that, except when the average quality of the attributes is high (the Bernoulli model with $p = 0.8$) or when the alternatives exhibit strong positive interattribute correlation (the PIAC model with $\rho = 0.5$), $P_S$ decays steeply with $k$ and has small values for large $k$. In fact, for large $k$, $P_S$ is never high. Thus, simple dominance cannot explain the observed good performance of certain simple heuristics.

For a fixed number of attributes, $P_S$ first decreases with the number of alternatives $m$ up to a certain value of $m$, $m^*$, beyond which it increases with $m$. This pattern can be explained by the three effects that a new alternative may have on an existing set of alternatives. The first possibility is that the new alternative simply dominates all others, making the new $S_k$ nonempty irrespective of whether or not it was empty before. Second, the new alternative may be simply dominated by some alternative, leaving $S_k$ unchanged. The third possibility is that the additional alternative neither simply dominates all others nor is simply dominated by any alternative, making the new $S_k$ empty if it was nonempty before. The first effect would force an increase with $m$ of $P_S$, whereas the third effect would force a decrease. As $m$ increases, the probability that the new alternative neither simply dominates all others nor is simply dominated by any other alternative becomes small, and for large enough $m$ the third effect is negligible and $P_S$ increases with $m$ as a result of the first effect. As shown in Theorem 6, $P_S$, and hence $P_C$, tends to one as $m$ tends to infinity. The $m^*$ turning point seems to increase as the number of attributes $k$ increases and as the average quality of the attributes decreases ($p$ gets smaller).

The low-to-high and the high-to-low patterns are identical. In hindsight, this is trivial because in simple dominance, nothing depends on the ordering of the attributes. Hence, any permutation across attributes would produce the same values of $P_S$. The probabilities in these two patterns are contained between the probabilities of the Bernoulli model with $p = 0.2$ and $p = 0.8$, which is intuitive. With respect to the base-case scenario, $P_S$ is larger for small $m$ and is similar for large $m$. As a result, the turning point $m^*$ increases.

### 3.4. Probability of Having Cumulative Dominance

Our goal in this subsection is to compute the probability of the presence of cumulative dominance for our probabilistic models. The computation of $P_C$ is at first glance significantly harder than the computation of $P_S$. Assuming that

cumulative dominance holds up to attribute $r$, one needs to keep track of all the attribute values up to attribute $r$ to know if cumulative dominance carries to attribute $r + 1$. This makes inefficient the use of Markov chain approaches to derive recursion-based expressions for $P_S$.

We have taken another approach that exploits the binary nature of our environment, based on reduced ordered binary decision diagrams (ROBDDs). We will first provide a brief overview of ROBDDs and explain how ROBDDs can be used to compute the probability that any given Boolean function of independent Bernoulli random variables takes value one. We will provide an illustrative small example. Next, we will explain with some detail how we used ROBDDs to compute $P_C$ for the ZIAC and PIAC models. Readers not interested in the ROBDD technique can skip directly to the results, shown in Figure 4 and discussed in the last three paragraphs of this subsection.

Of course, we could have used Monte Carlo simulation to estimate $P_S$ and $P_C$. However, such an approach would be more costly in terms of CPU times than our ROBDD-based approach if one wants $P_S$ and $P_C$ to be estimated with high accuracy. In addition, a numerically exact solution, if available, is always better than a simulation. The ROBDDs allow us to calculate $P_S$ and $P_C$ for sufficiently large values of $m$ and $k$ and support the main message of the paper: Cumulative dominance-compliant heuristics and fully cumulative dominance-compliant heuristics perform well in many cases.

An ROBDD (Bryant 1986) is a directed acyclic graph having a single root node and two terminal nodes (leaves), one labeled zero and another labeled one. An ROBDD is able to represent any arbitrary binary function of binary variables, i.e., any Boolean function. Each nonterminal node has two outcoming edges: a 0-edge and a 1-edge. ROBDDs are called reduced because each node represents a different Boolean function. The root node represents the given Boolean function. The terminal node 0 represents the "0" Boolean function. The terminal node 1 represents the "1" Boolean function. Each nonterminal node has a binary variable associated with it. Consider a node $n$ having associated with it variable $x$. Let $n_0$ and $n_1$ be the outcoming 0-edge and 1-edge nodes of $n$, respectively. The nodes $n_0$ and $n_1$ represent the Boolean functions obtained from the Boolean function represented by $n$ by setting $x$ to zero and one, respectively. ROBDDs are called ordered because they depend on the ordering of the binary variables. To build the ROBDD of a Boolean function $F(x_1, x_2, \ldots, x_n)$, a description of $F()$ in terms of $x_1, x_2, \ldots, x_n$ using basic Boolean operators like NOT ($\neg$), AND ($\wedge$), and OR ($\vee$) is required. That description can be looked at as a single output combinational circuit (Ercegovac and Lang 1985). The ROBDD of the function can then be obtained by traversing depth first the combinational circuit starting at the output. When returning from the visit of every noninput node, the ROBDD of the node can be obtained by combining the ROBDDs of the inputs using well-known procedures (Bryant 1986). Given the ROBDD of a function of $n$

**Figure 2.** $P_S$ as a function of $m$ (horizontal axis) and $k$ for the base-case scenario and the six variations.
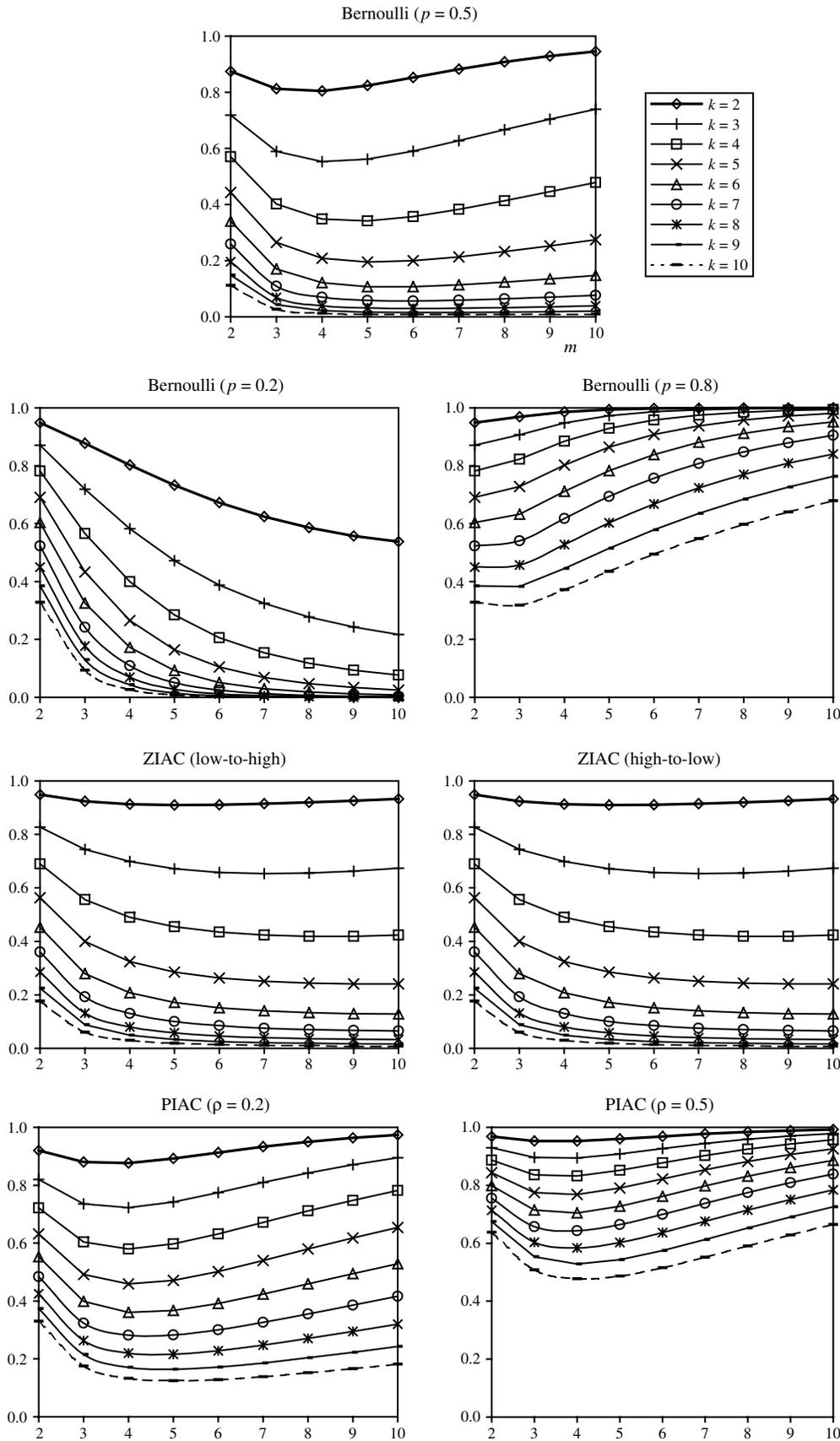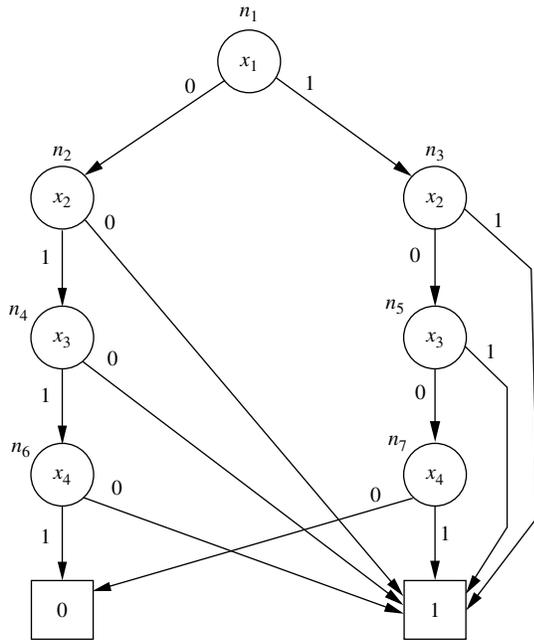
**Figure 3.** ROBDD of a function $F(x_1, x_2, x_3, x_4)$ for the variable ordering $x_1, x_2, x_3, x_4$.



independent Bernoulli random variables $F(x_1, x_2, \ldots, x_n)$, computing $P[F(x_1, x_2, \ldots, x_n) = 1]$ is a trivial task. It suffices to perform a depth-first traversal of the ROBDD starting at the root node. When returning from the visit of every nonterminal node $n$ having associated with it variable $x$, we obtain the probability that the function represented by $n$ is equal to one by multiplying $P[x = 0]$ by the probability that the function represented by the 0-edge node is equal to one, multiplying $P[x = 1]$ by the probability that the function represented by the 1-edge node is equal to one, and adding up those partial results. Of course, this is possible because the functions represented by the 0-edge and 1-edge nodes do not depend on $x$.

As an illustration, Figure 3 depicts the ROBDD of a certain Boolean function $F(x_1, x_2, x_3, x_4) = \neg[(\neg x_1 \wedge x_2 \wedge x_3 \wedge x_4) \vee (x_1 \wedge \neg x_2 \wedge \neg x_3 \wedge \neg x_4)]$ for the variable ordering $x_1, x_2, x_3, x_4$. The root node is node $n_1$. Node $n_4$, for example, represents the Boolean function of $x_3$ and $x_4$ obtained from $F(x_1, x_2, x_3, x_4)$ by setting the variables $x_1$ to zero and $x_2$ to one. The problem is to calculate the probability that $F$ takes value one, given that the instances of $x_i$ are independent Bernoulli random variables with parameter $p$. Assuming that preference is given to the 0-edges over the 1-edges during the depth-first traversal of the ROBDD, nonterminal nodes would be visited in the order $n_1, n_2, n_4, n_6, n_3, n_5, n_7$. The probabilities, $q_m$, $1 \leqslant m \leqslant 7$, that the functions represented by the nonterminal nodes $n_m$ have value one would be computed in the ordering and with the results:

$$q_6 = 1 - p,$$
$$q_4 = (1 - p) + p \times q_6 = 1 - p^2,$$
$$q_2 = (1 - p) + p \times q_4 = 1 - p^3,$$

$$q_7 = p,$$
$$q_5 = (1 - p) \times q_7 + p = 2p - p^2,$$
$$q_3 = (1 - p) \times q_5 + p = 3p - 3p^2 + p^3,$$
$$q_1 = (1 - p) \times q_2 + p \times q_3 = 1 - p + 3p^2 - 4p^3 + 2p^4.$$

In our particular application, computation of $P_C$ using ROBDDs was performed by noting that $P_C = P[C_k \neq \varnothing] = P[\mathbf{1}_{\{C_k \neq \varnothing\}} = 1]$, $\mathbf{1}_A$ denoting the indicator function of event $A$, and that the binary random variables underlying the ZIAC and PIAC models are independent. For the ZIAC model, the independent Bernoulli random variables are $x_{i,s}$, $1 \leqslant i \leqslant m$, $1 \leqslant s \leqslant k$, and the combinational circuit was built based on

$$\mathbf{1}_{\{C_k \neq \varnothing\}} = \bigvee_{i=1}^{m} \bigwedge_{\substack{j=1 \\ j \neq i}}^{m} \bigwedge_{s=1}^{k} \mathbf{1}_{\{X_{i,s} \geqslant X_{j,s}\}}.$$

For $s = 1$, the logic used to generate $\mathbf{1}_{\{X_{i,1} \geqslant X_{j,1}\}}$, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant m$, $j \neq i$ was

$$\mathbf{1}_{\{X_{i,1} \geqslant X_{j,1}\}} = \mathbf{1}_{\{x_{i,1} \geqslant x_{j,1}\}} = x_{i,1} \vee \neg x_{j,1}.$$

For generating $\mathbf{1}_{\{X_{i,s} \geqslant X_{j,s}\}}$, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant m$, $j \neq i$, $2 \leqslant s \leqslant k$, we used $m(k-1)$ adders $A_{i,s}$, $1 \leqslant i \leqslant m$, $2 \leqslant s \leqslant k$ and $m(m-1)(k-1)$ comparators $C_{i,j,s}$, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant m$, $j \neq i$, $2 \leqslant s \leqslant k$. Adder $A_{i,s}$ was a specialized carry propagate adder having as inputs the $b(s-1)$-bit codeword $X_{i,s-1}^{b(s-1)}, X_{i,s-1}^{b(s-1)-1}, \ldots, X_{i,s-1}^1$, $b(s-1) = \lceil \log_2(s) \rceil$ representing in binary code $X_{i,s-1} = \sum_{t=1}^{s-1} x_{i,t}$ and the $b(s-1)$-bit codeword $00 \cdots 0 x_{i,s}$, and as output, the $b(s)$-bit codeword $X_{i,s}^{b(s)}, X_{i,s}^{b(s)-1}, \ldots, X_{i,s}^1$, $b(s) = \lceil \log_2(s+1) \rceil$ representing in binary code $X_{i,s} = \sum_{t=1}^{s} x_{i,t}$, and was implemented with NOT, AND, and OR gates. Comparator $C_{i,j,s}$ was an iterative bit-sliced comparator having as inputs the $b(s)$-bit codewords $X_{i,s}^{b(s)}, X_{i,s}^{b(s)-1}, \ldots, X_{i,s}^1$ and $X_{j,s}^{b(s)}, X_{j,s}^{b(s)-1}, \ldots, X_{j,s}^1$, $b(s) = \lceil \log_2(s+1) \rceil$ and as output $\mathbf{1}_{\{X_{i,s} \geqslant X_{j,s}\}}$, and was also implemented with NOT, AND, and OR gates.

For the PIAC model, the independent Bernoulli random variables to be considered are $z_i$, $1 \leqslant i \leqslant m$, $y_{i,s}^l$, $1 \leqslant i \leqslant m$, $1 \leqslant s \leqslant k$, and $y_{i,s}^h$, $1 \leqslant i \leqslant m$, $1 \leqslant s \leqslant k$, and the combinational circuit yielding $\mathbf{1}_{\{C_k \neq \varnothing\}}$ was built as for the ZIAC model adding the logic to obtain $x_{i,s}$, $1 \leqslant i \leqslant m$, $1 \leqslant s \leqslant k$,

$$x_{i,s} = \neg z_i \wedge y_{i,s}^l \vee z_i \wedge y_{i,s}^h.$$

There are packages that simplify the use of ROBDDs. Perhaps the most well known currently is the CU Decision Diagram Package (Somenzi 2005), which we used. This package uses a slight variant of ROBDDs, called ROBDDs with complement 0-edges (Brace et al. 1990). They are like standard ROBDDs except for the facts that the root node may represent either the given Boolean function or its complement and that nonterminal nodes may

have either an outcoming 0-edge and an outcoming 1-edge or an outcoming complement 0-edge and an outcoming 1-edge. An outcoming complement 0-edge of a node $n$ having associated with it variable $x$ leads to a node that represents the *complement* of the Boolean function obtained from the Boolean function represented by $n$ by setting $x$ to zero. ROBDDs with complement 0-edges have, potentially, a smaller number of nodes than standard ROBDDs. Adapting the procedure for computing the probability that a Boolean function of Bernoulli random variables is equal to one to ROBDDs with complement 0-edges is trivial. First, if the root node represents the complement of the given Boolean function, it suffices to note that the probability that the given Boolean function has value one is one minus the probability that the Boolean function represented by the root node is equal to one. Second, let $n$ be the node from which some complement 0-edge comes out, and let $x$ be the variable associated with $n$; then, the probability that the Boolean function obtained from the Boolean function represented by $n$ by setting $x$ to zero is one minus the probability that the Boolean function represented by the complement 0-edge node is equal to one.

Use of computer memory tends to be the main limitation when using ROBDDs. Memory consumption is determined by the peak number of reserved nodes during the ROBDD construction. The CU Decision Diagram Package includes features to reduce that peak number as much as possible. We use these features as efficiently as possible to reduce the number of reserved nodes, while maintaining small CPU times. The size of ROBDDs depends on the variable ordering. The variable ordering is typically chosen using heuristics based on the description of the function. We decided to use the topology heuristic (Nikolskaia et al. 1998). Using this heuristic, we were able to compute in small CPU times and with modest memory consumption the probability $P_C$ for both probabilistic models for values of $m$ and $k$ as large as 10. The peak numbers of reserved nodes increased with both $m$ and $k$ and were a bit larger for the PIAC model. For $m = 10$ and $k = 10$, the peak numbers of reserved nodes was 2,325,051 for the ZIAC model and 2,449,735 for the PIAC model. The total CPU time to calculate $P_C$ for $m = 10$ and $k = 10$ in a modest computer platform was 79 seconds for the ZIAC model, and 381 seconds for the PIAC model.

Figure 4 plots $P_C$ for values of $k$ ranging from 3 to 10 and values of $m$ ranging from 2 to 10 (for binary attributes, $r^* \geqslant 2$ guarantees $P_C = 1$ for $k = 2$). The seven graphs correspond to the base-case scenario and the six variations. In all cases, $P_C$ is significantly larger than $P_S$ (Figure 2). As for $P_S$, for a fixed number of alternatives $m$, $P_C$ decreases with $k$. However, $P_C$ decreases at a much slower rate than $P_S$. For fixed $k$, there exists a turning point, $m^*$, for $m$ before which $P_C$ decays with $m$ and beyond which $P_C$ increases with $m$. The explanation for these turning points is a replica of the reasons given in the case of simple dominance, but in terms of cumulative dominance. In fact, as shown in Theorem 6, for fixed $k$, $P_C \to 1$ as $m \to \infty$.

In the high-to-low pattern, $P_C$ improves over the base-case scenario. Intuitively, if the probability of high values for the tail attributes is low, then the cumulative dominance acquired up to some attribute $r < k$ will be easily preserved all the way until attribute $k$. In contrast, the performance deteriorates in the low-to-high pattern. The intuition for this last result is similar. There is a greater chance that the best alternative has high values in the tail attributes, but fails to cumulative dominate the others because it had low values in the first attributes. This also implies that in the low-to-high pattern, the expected loss will be larger than in the high-to-low pattern. We will verify this claim in the next subsection.

For fixed $k$ and $m$, the values of $P_C$ improve with the average quality of the attributes (Bernoulli model, higher $p$), with a high-to-low pattern for average attribute qualities (ZIAC model, high-to-low pattern) and with positive interattribute correlation (PIAC model, higher $\rho$). $P_C$ is close to one in many cases in which $P_S$ is not close to one. In particular, $P_C$ is close to one, even when the number of attributes is large, when either the average quality of the attributes is high or when there exists strong positive interattribute correlation. It is also noteworthy that, in contrast with $P_S$ and contrary to intuition, $P_C$ has significant values even when the attributes are of poor average quality, attributes exhibit a low-to-high average quality pattern, or attributes are uncorrelated. Summarizing, the presence of cumulative dominance predicts good efficiency for cumulative dominance-compliant heuristics.

### 3.5. Upper Bound for the Expected Loss of Fully Cumulative Dominance-Compliant Heuristics

The probability that a heuristic chooses a best alternative is an important metric of performance. A probability close to one guarantees an excellent performance of heuristics that comply with cumulative dominance. However, if we are in an environment with low levels of cumulative dominance, then it is important to know the average or expected loss of the heuristic. Arguably, the loss is a superior measure of performance: it takes the value zero when the heuristic chooses the best alternative, and otherwise, positive values measuring the utility loss. In many cases, we would be content with an inferior alternative if we knew that its utility was reasonably close to the utility of the best alternative. In this section, we derive an upper bound for the expected loss of any fully cumulative compliant heuristic.

In the context of a probability model, $L$, as defined in (3), is a random variable. Our goal is to provide an upper bound, $E[L]_{\text{ub}}$, for its expectation. Using Theorem 5, all that is required is to take expectations and to condition on the value of $r^*$:

$$E[L] \leqslant E\left[\frac{k - r^*}{k}\right] = \sum_{r=1}^{k} P[r^* = r]\frac{k - r}{k}$$

$$= \sum_{r=1}^{k-1} P[r^* = r]\frac{k - r}{k}.$$

**Figure 4.**    $P_C$ as a function of *m* and *k* for the base-case scenario and the six variations.
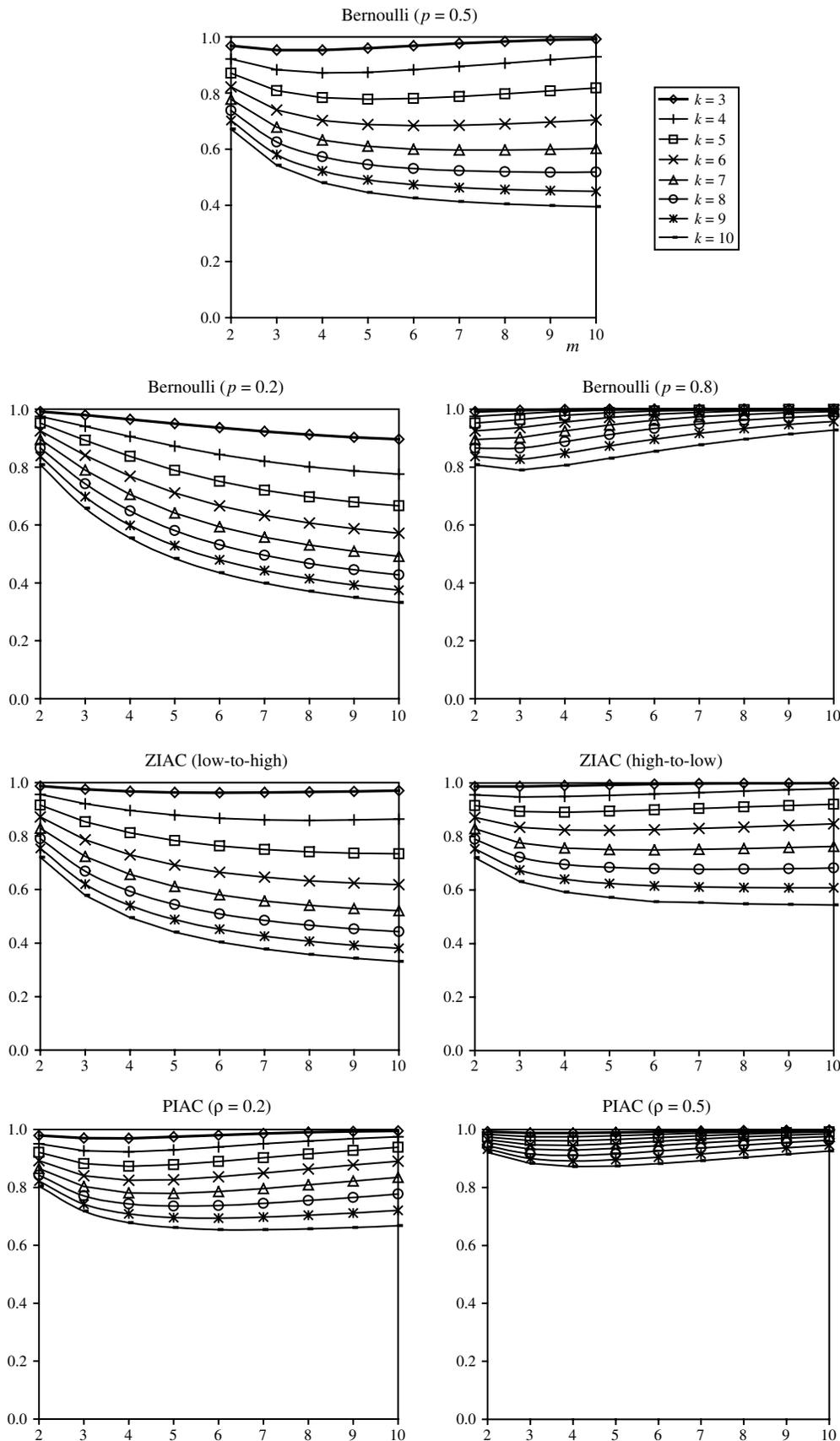
In the binary attribute case, $P[r^* = 1] = 0$. To calculate $P[r^* = r]$, $2 \leqslant r \leqslant k - 1$ for the two considered probabilistic models, let $Q(r) = P[r^* \geqslant r]$. We have

$$P[r^* = r] = Q(r) - Q(r+1), \quad 2 \leqslant r \leqslant k - 1.$$

Because $r^* \geqslant 2$, $Q(2) = 1$. Noting that $Q(r) = P[C_r \neq \varnothing]$, the values of $Q(r)$, $3 \leqslant r \leqslant k$ can be obtained by using the ROBDD approaches described in §3.4 to compute $P_C = P[C_k \neq \varnothing]$ with the index $k$ replaced by the index $r$.

Figure 5 plots $E[L]_{\text{ub}}$ for values of $k$ ranging from 3 to 10 and values of $m$ ranging from 2 to 10, for the base-case scenario and the six variations. For fixed $m$, $E[L]_{\text{ub}}$ increases with $k$, but slowly. For fixed $k$, there exists a turning point $m^*$ before which $E[L]_{\text{ub}}$ increases with $m$ and beyond which $E[L]_{\text{ub}}$ decreases with $m$. The fact that, for fixed $k$, the probability that some alternative will have all its attribute values equal to one tends to one as $m \to \infty$ ensures that for fixed $k$, $E[L]_{\text{ub}}$ will tend to zero as $m \to \infty$. Not surprisingly, the value of $E[L]_{\text{ub}}$ is very small when the attributes have good average quality (Bernoulli model, $p = 0.8$) and when there exists strong positive interattribute correlation (PIAC model, $\rho = 0.5$). More importantly, there are cases (see Figure 4) in which $P_C$ is not high and $E[L]_{\text{ub}}$ is reasonably small. In particular, $E[L]_{\text{ub}}$ is reasonably small, even when the number of attributes is large in the presence of a high-to-low pattern in the average attribute quality (ZIAC model, high-to-low pattern) and when there is moderate positive interattribute correlation (PIAC model, $\rho = 0.2$). Thus, the predicted good performance of cumulative dominance-compliant heuristics is made stronger for fully cumulative dominance-compliant heuristics.

Although $E[L]_{\text{ub}}$ is a reasonably good upper bound and applies to all heuristics that fully comply with cumulative dominance, much better upper bounds for $E[L]_{\text{ub}}$ can be obtained for the particular case of *DEBA*. Two of the authors have worked out improved upper bounds for the particular case of *DEBA* (Carrasco and Baucells 2008). The improved upper bounds reduce the values presented here roughly by half.

## 4. Cumulative Dominance and Heuristics Performance

### 4.1. DEBA

As described above, *DEBA* is a lexicographic heuristic that can be formalized as follows. Consider $m$ alternatives, each with $k$ attributes taking real values (the binary case is a particular case). *DEBA* proceeds in steps. Let $A_r$, $1 \leqslant r \leqslant k$, be the subsets of alternatives that survive the selection steps of *DEBA*. The subset $A_1$ includes the alternatives with largest values of $x_{i,1}$. $A_2$ includes the alternatives in $A_1$ with largest $x_{i,2}$, etc. Obviously, $A_1 \supset \cdots \supset A_k$. The set $A_k$ is guaranteed to be nonempty and all alternatives in $A_k$ have the

same profile: *DEBA* finalizes by choosing at random from the alternatives in $A_k \neq \varnothing$.

We now argue that *DEBA* fully complies with cumulative dominance. This can be seen informally as follows. For some $r$, assume that $C_r$ is nonempty, and consider some alternative $j$ not in $C_r$. Alternative $j$ is not in $C_r$ for a reason: it is the case that for some attribute $t \leqslant r$, $x_{j,t}$ was smaller than $x_{i,t}$, $i \in C_k$. Hence, $j$ cannot survive *DEBA*'s elimination step $t$.

**THEOREM 9.** DEBA *is fully cumulative dominance compliant.*

**PROOF.** We will show that $A_r = C_r$, $1 \leqslant r \leqslant r^*$. The result then follows because *DEBA* chooses an alternative from $A_k$, and $A_k \subset A_{r^*} = C_{r^*}$. That $C_r \subset A_r$, $1 \leqslant r \leqslant k$, can be seen by induction on $r$. Obviously, $C_1 = A_1$. Assume that the result holds for $r = s$ and consider the case $r = s + 1$. Let $i \in C_{s+1}$. We have $X_{i,s+1} \geqslant X_{j,s+1}$, $1 \leqslant j \leqslant m$, $j \neq i$. Because $C_{s+1} \subset C_s$, by the induction hypothesis $i \in A_s$. Assume that $i \notin A_{s+1}$. Then, there exists an alternative $l \in A_{s+1}$, $l \neq i$, with $x_{l,s+1} > x_{i,s+1}$ and $x_{l,u} = x_{i,u}$, $1 \leqslant u \leqslant s$. However, this implies $X_{i,s+1} < X_{l,s+1}$ and, therefore, $i \notin C_{s+1}$, a contradiction. That $A_r \subset C_r$ for all $r$, $1 \leqslant r \leqslant r^*$ can be seen by contradiction. Take some $r$, $1 \leqslant r \leqslant r^*$, and an alternative $i$ such that $i \in A_r$ and $i \notin C_r$. Because all alternatives in $A_r$ are identical up to attribute $r$, this would imply $A_r \cap C_r = \varnothing$, which by $C_r \subset A_r$ implies $C_r = \varnothing$, a contradiction. Thus, $A_r = C_r$ for all $r$, $1 \leqslant r \leqslant r^*$. $\square$

As a procedure, *DEBA* generalizes—to more than two alternatives—the lexicographic binary-choice model Take-The-Best (*TTB*) proposed by Gigerenzer and Goldstein (1996). There is a small difference, however: in *TTB*, the attributes are ordered by their validities (ability to predict the criterion). The validities are computed using a database of previous instances of alternatives, whereas in *DEBA* the ordering of the attributes is assumed to be known.

Using simulations or particular data sets of alternatives, several studies have shown *TTB* to be effective in relation to alternative simple-decision heuristics (Gigerenzer and Goldstein 1996, Czerlinski et al. 1999, Martignon and Hoffrage 1999, 2002) as well as having desirable properties for both binary and multivariate choice (Hogarth and Karelaia 2006a, Katsikopoulos and Martignon 2006, Katsikopoulos and Fasolo 2006). Even when attributes are continuous variables, the model can be quite effective under some circumstances (Gigerenzer and Todd 1999, Hogarth and Karelaia 2005a). Most of these studies are restricted to the case of two or three alternatives. Finally, there is empirical evidence that people do sometimes use *DEBA*-like strategies in decision making (Bröder 2000, Newell and Shanks 2003, Newell et al. 2003).

The results in §§3.4 and 3.5 argue that the performance of *DEBA* can be explained by the high presence and persistence of cumulative dominance. This is not the only explanation: it is well known that if the weights are non-

**Figure 5.** $E[L]_{ub}$ as a function of $m$ and $k$ for the base-case scenario and the six variations.
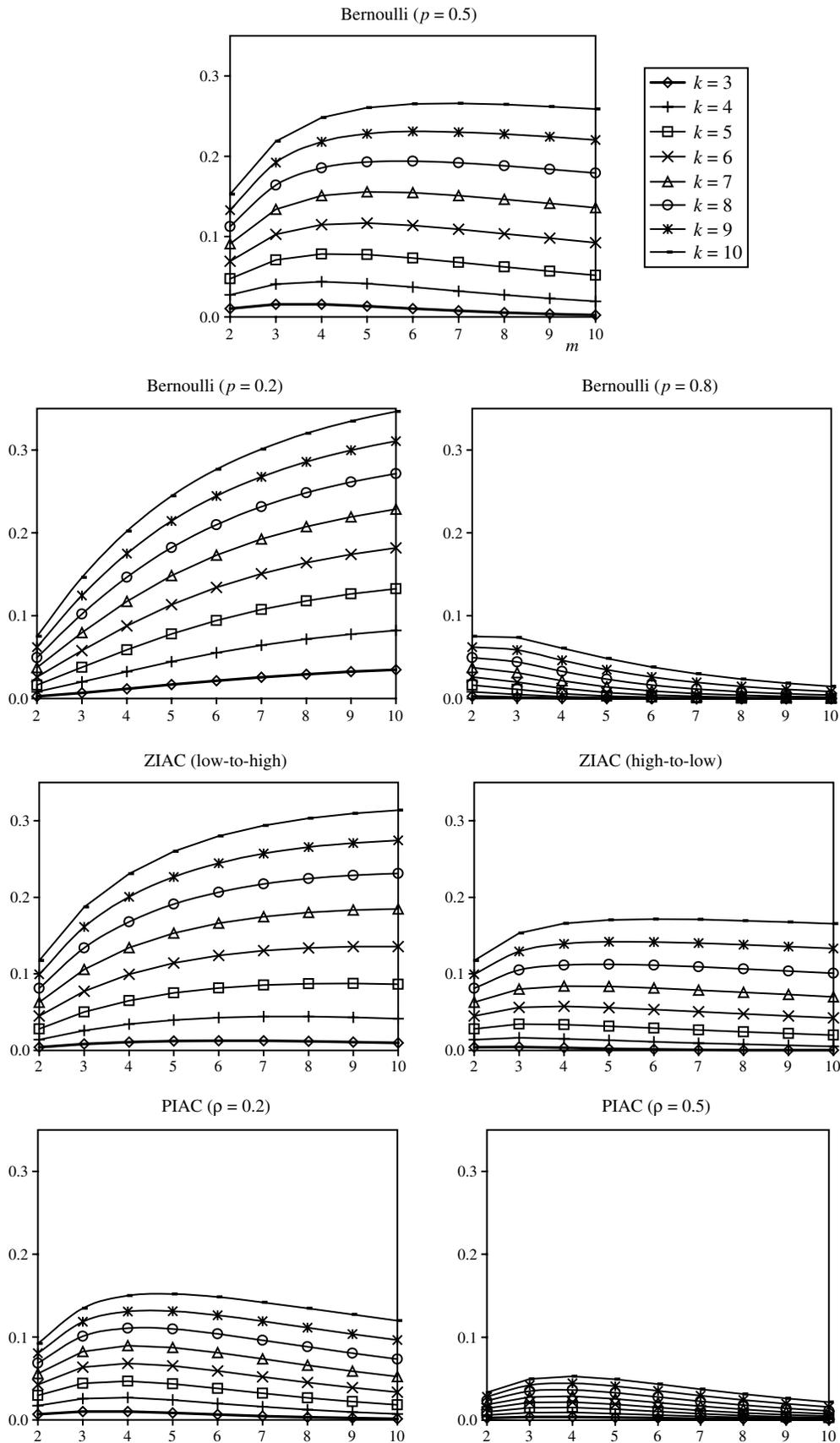
compensatory ($w_r \geqslant \sum_{s=r+1}^{k} w_s$, $1 \leqslant r \leqslant k - 2$), then both *TTB* and *DEBA* choose the best alternative with probability one (Martignon and Hoffrage 1999, 2002). In other words, the probability lower bound $P_C$ and the upper bound for $E[L]$ are pessimistic because they hold for the most pessimistic set of weights. In summary, *DEBA* works very well either in the presence of cumulative dominance or when weights are noncompensatory.

We do not know of any other heuristic discussed in the literature that is fully cumulative dominance compliant. We now introduce another fully compliant heuristic different from *DEBA*. We call it *CDS/EW* (Cumulative-Dominance-Selection/Equal-Weights). *CDS/EW* first considers the alternatives in $C_{r^*}$ and, from those, selects the one with largest attribute sum $X_{i,k}$, breaking ties at random. Although *CDS/EW* is more onerous to apply than *DEBA*, *CDS/EW* is intuitively appealing: first, it maximizes with certainty the part of the utility corresponding to attributes $1, 2, \ldots, r^*$, and then takes a more global view than *DEBA* to try to maximize the utility corresponding to the attributes $r^* + 1, \ldots, k$. This strategy might be advantageous if $r^*$ is not close to $k$.

## 4.2. Heuristics Based on Equal Weights

Equal weights-based heuristics (*EW*) have been studied, among others, by Dawes and Corrigan (1974), Einhorn and Hogarth (1975), and Payne et al. (1993). In the binary attribute case, those heuristics frequently encounter tied alternatives and require an additional step to break ties. One such possibility is *EW/RAN*, which chooses at random between the alternatives with the highest attribute sum. While *EW/RAN* complies with simple dominance, it does not comply with cumulative dominance. To see this, consider the case of two alternatives with profiles $x_1 = (1, 0, 1, 1)$ and $x_2 = (0, 1, 1, 1)$. The profile sum is three in both cases, so that *EW/RAN* may end up choosing alternative 2, which is cumulative dominated by alternative 1. Not surprisingly, the performance of *EW/RAN* decays quite rapidly with $m$ and $k$ (Hogarth and Karelaia 2006a, 2005b).

Another possibility is to break ties using *DEBA*. The combination of *EW* and *DEBA* gives rise to the *EWn/DEBA* family of heuristics. For fixed value of $n$, $1 \leqslant n \leqslant k$, *EWn/DEBA* first considers the alternatives with the largest values of $X_{i,n}$, and uses *DEBA* to choose from this subset. The *EWn/DEBA* class contains two important particular cases. At one extreme ($n = k$), we get *EW/DEBA*, which first chooses the alternatives with largest $X_{i,k}$ and breaks ties using *DEBA*. In the binary attribute case, *EW/DEBA* is quite intuitive. It chooses the alternative with a higher number of "pluses," and then uses *DEBA* to break ties. At the other extreme ($n = 1$, $n = 1, 2$ if attributes are binary), we get *DEBA*. We now argue that *EWn/DEBA* complies with cumulative dominance.

THEOREM 10. *EWn/DEBA*, $1 \leqslant n \leqslant k$, *is cumulative dominance compliant.*

PROOF. Assume that $C_k \neq \varnothing$. Note that the alternatives in $C_k$ cumulative dominate all others, and hence have the largest values of $X_{i,r}$, $1 \leqslant r \leqslant k$. In particular, they have the largest values of $X_{i,n}$. Hence, the first phase of *EWn/DEBA* will not remove these from consideration. Of course, other alternatives with the same value of $X_{i,n}$ may remain. However, *DEBA* is used to break ties, and *DEBA* fully complies with cumulative dominance. Hence, those other alternatives will now be removed, and the heuristic will finally choose an alternative from $C_k$. □

Attractive as it seems, *EWn/DEBA* is not fully cumulative dominance compliant. Consider, for example, the case of two alternatives with profiles $x_1 = (1, 0, 0)$ and $x_2 = (0, 1, 1)$. In this case, $C_{r^*} = C_2 = \{1\}$, but *EW3/DEBA* chooses alternative 2. An important implication is that the upper bound on the loss given by Theorem 5 does not apply to *EWn/DEBA*. In this example, the loss guaranteed by any heuristic that fully complies with cumulative dominance is $\leqslant (k - r^*)/k = 1/3$. *DEBA* chooses alternative 1, and the loss in the most pessimistic weight scenario ($w_1 = w_2 = w_3 = 1/3$) is $L = U_2 - U_1 = 1/3$, in accordance with the upper bound. In contrast, *EW3/DEBA* chooses alternative 2, and for appropriate weights ($w_1 = 1 - 2\varepsilon$, $w_2 = w_3 = \varepsilon$, $\varepsilon \approx 0$), this choice yields a loss of $L = U_1 - U_2 = 1 - 4\varepsilon \approx 1$.

## 5. Final Remarks and Conclusions

Using the notion of cumulative dominance, we have justified, for the binary attribute case and for two probabilistic models, the observed effectiveness of the *DEBA* heuristic. The results obtained in the paper are applicable to any heuristic that complies or fully complies with cumulative dominance. We have provided existing and nonexisting examples of heuristics in those classes different from *DEBA*. Our results can be used to bound the performance of those heuristics independently of the particular values of the weights. Our computational procedures are feasible for quite large values of $m$ and $k$. Previous studies concerning the performance of *DEBA* and *EWn/DEBA* (Hogarth and Karelaia 2006a, 2005b) used simpler enumeration approaches and more modest values of $m$ and $k$ ($m$ up to five and $k$ up to five).

Our study is one more step in the direction of reducing the descriptive-prescriptive gap in multiattribute decision making. We have shown that *DEBA* and other related heuristics can be effective in the binary attribute setting in many cases. This strongly supports the insight that the key managerial skill is to identify and rank the most relevant attributes or factors. Efforts to specify exact values of weights and/or use information-intensive decision procedures may have minor returns and be justified only for a small fraction of decisions (Keeney 2004).

This work can be generalized in several directions. First, it would be interesting to analyze the impact of negative interattribute correlation. For example, alternatives

offered in the marketplace usually exhibit negative correlation between price and quality. However, negative correlation between one attribute and the rest induces positive correlation between the latter. Hence, besides the magnitude of the negative correlation, the model has to specify the position in the ordering of the negatively correlated attribute. In summary, the study of negative correlation, or of general correlation attribute structures, cannot be done in the same parsimonious way as the positive correlation case, and a complete treatment is left for further research.

The formality of our results could be applied to other, very different settings. As suggested by one of the reviewers, another scenario to justify the model would be the attributes being the approval or disapproval of members of a committee. An alternative having a profile of 10011 means that committee member 1 favors it, committee member 2 opposes it, and so on. The assumption is that some members are more influential than others (e.g., more senior, more expert, with a greater stake in the choice). This interpretation is attractive because it avoids attaching specific weights to group members.

It is important to emphasize that the properties of (full) cumulative dominance compliance also hold for general discrete and continuous attributes. Simulation results indicate that in the discrete case, the presence of cumulative dominance deteriorates as the number of attribute levels increases (Kirkwood and Corner 1993). However, the case of equally likely attribute levels might not be representative—hence, the importance of studying other environments such as normally distributed continuous attribute values, possibly correlated. Indeed, the analytic framework recently developed by Hogarth and Karelaia (2006b) provides a means for beginning to study this issue. Furthermore, it would be illuminating to investigate the level of cumulative dominance in naturally occurring decision environments.

Even if there is no cumulatively dominant alternative, some alternatives can be eliminated because they are cumulatively dominated by others. Hence, cumulative dominance establishes a partial order. This suggests questions, left for future research, such as: What is the expected number of alternatives remaining after applying cumulative dominance considerations?, or, What is the probability that cumulative dominance produces a complete ordering?

## 6. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.pubs.informs.org/.

## Acknowledgments

## References

Barron, F. H., B. E. Barrett. 1996. Decision quality using ranked attribute weights. *Management Sci.* **42**(11) 1515–1523.

Baucells, M., R. Sarin. 2003. Group decisions with multiple criteria. *Management Sci.* **49**(8) 1105–1118.

Beardon, A. F. 1979. *Complex Analysis. The Argument Principle in Analysis and Topology*. John Wiley & Sons, New York.

Brace, K. S., R. L. Rudell, R. E. Bryant. 1990. Efficient implementation of a BDD package. *Proc. 27th ACM/IEEE Design Automation Conf.*, 40–57.

Bröder, A. 2000. Assessing the empirical validity of the "Take-the-Best" heuristic as a model of human probabilistic inference. *J. Experiment. Psych.: Learn., Memory, Cognition* **26**(5) 1332–1346.

Bryant, R. E. 1986. Graph-based algorithms for Boolean function manipulation. *IEEE Trans. Comput.* **C-35**(8) 677–691.

Butler, J. C., J. Jia, J. S. Dyer. 1997. Simulation techniques for the sensitivity analysis of multi-criteria decision models. *Eur. J. Oper. Res.* **103** 531–545.

Carrasco, J. A., M. Baucells. 2008. Tight upper bounds for the expected loss of lexicographic heuristics in binary multiattribute choice. *Math. Soc. Sci.* **55** 156–189.

Czerlinski, J., G. Gigerenzer, D. G. Goldstein. 1999. How good are simple heuristics? G. Gigerenzer, P. M. Todd (ABC Research Group), eds. *Simple Heuristics That Make Us Smart*. Oxford University Press, New York, 97–118.

Dawes, R. M., B. Corrigan. 1974. Linear models in decision making. *Psych. Bull.* **81**(2) 95–106.

Durret, R. 2004. *Probability: Theory and Examples*, 3rd ed. Duxbury Press, Belmont, CA.

Dyer, J. S., R. K. Sarin. 1979. Measurable multiattribute value functions. *Oper. Res.* **27** 810–822.

Einhorn, H. J., R. M. Hogarth. 1975. Unit weighting schemes for decision making. *Organ. Behav. Human Performance* **13**(2) 171–192.

Ercegovac, M. D., T. Lang. 1985. *Digital Systems and Hardware/Firmware Algorithms*. John Wiley & Sons, New York.

Gigerenzer, G., D. Goldstein. 1996. Reasoning the fast and frugal way: Models of bounded rationality. *Psych. Rev.* **103**(4) 650–669.

Gigerenzer, G., P. M. Todd (ABC Research Group), eds. 1999. *Simple Heuristics That Make Us Smart*. Oxford University Press, New York.

Hogarth, R. M., N. Karelaia. 2005a. Ignoring information in binary choice with continuous variables: When is less more? *J. Math. Psych.* **49**(2) 115–124.

Hogarth, R. M., N. Karelaia. 2005b. Simple models for multiattribute choice with many alternatives: When it does and does not pay to face trade-offs with binary attributes. *Management Sci.* **51**(12) 1860–1872.

Hogarth, R. M., N. Karelaia. 2006a. "Take-the-Best" and other simple strategies: Why and when they work "well" with binary cues. *Theory and Decision* **61** 205–249.

Hogarth, R. M., N. Karelaia. 2006b. Regions of rationality: Maps for bounded agents. *Decision Anal.* **3**(3) 124–144.

Hsee, C. K., Y. Rottenstreich. 2004. Music, pandas, and muggers: On the affective psychology of value. *J. Experiment. Psych.* **133**(1) 23–30.

Katsikopoulos, K. V., B. Fasolo. 2006. New tools for decision analysts. *IEEE Trans. Systems, Man, Cybernetics Part A: Systems and Humans* **36**(5) 960–967.

Katsikopoulos, K. V., L. Martignon. 2006. Naive heuristics for paired comparisons: Some results on their relative accuracy. *J. Math. Psych.* **50** 488–494.

Keeney, R. L. 2004. Making better decision makers. *Decision Anal.* **1**(4) 193–204.

Keeney, R. L., H. Raiffa. 1976. *Decisions with Multiple Objectives: Preference and Value Tradeoffs*. John Wiley, New York.

Kirkwood, C. W., J. L. Corner. 1993. The effectiveness of partial information about attribute weights for ranking alternatives in multiattribute decision making. *Organ. Behav. Human Decision Processes* **54** 456–476.

Kirkwood, C. W., R. K. Sarin. 1985. Ranking with partial information: A method and an application. *Oper. Res.* **33**(1) 193–204.

Martignon, L., U. Hoffrage. 1999. Why does one-reason decision making work? A case study in ecological rationality. G. Gigerenzer, P. M. Todd (ABC Research Group), eds. *Simple Heuristics That Make Us Smart*. Oxford University Press, New York, 119–140.

Martignon, L., U. Hoffrage. 2002. Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision* **52** 29–51.

Newell, B. R., D. R. Shanks. 2003. Take the best or look at the rest? Factors influencing one-reason decision making. *J. Experiment. Psych.: Learn., Memory, Cognition* **29**(1) 53–65.

Newell, B. R., N. J. Weston, D. R. Shanks. 2003. Empirical tests of a fast-and-frugal heuristic: Not everyone takes-the-best. *Organ. Behav. Human Decision Processes* **91**(1) 82–96.

Nikolskaia, M., A. Rauzy, D. J. Sherman. 1998. Almana: A BDD minimization tool integrating heuristic and rewriting methods. *Formal Methods in Comput.-Aided Design, 2nd Internat. Conf.*, Springer-Verlag LNCS, Vol. 1522, 100–114.

Payne, J. W., J. R. Bettman, E. J. Johnson. 1993. *The Adaptive Decision Maker*. Cambridge University Press, New York.

Rao, J. S., M. Sobel. 1980. Incomplete Dirichlet integrals with applications to ordered uniform spacings. *J. Multivariate Anal.* **10** 603–610.

Somenzi, F. 2005. CUDD: CU decision diagram package release 2.4.1. http://vlsi.colorado.edu/~fabio/CUDD/cuddIntro.html.

Tversky, A. 1972. Elimination by aspects: A theory of choice. *Psych. Rev.* **79**(4) 281–299.