Regions of rationality: Maps for bounded agents

Robin M. Hogarth[a*] & Natalia Karelaia[b]

ICREA[a] & Universitat Pompeu Fabra[a], Barcelona

HEC, Lausanne,[b] Switzerland

March 2006

∗ The authors are grateful for the helpful comments of the editors and anonymous reviewers as well as for feedback received at presentations made at the Center for Decision Research at the University of Chicago, Universitat Pompeu Fabra, Carnegie-Mellon University, the Haas School of Business at the University of California, Berkeley, the Cogeco Conference in Sofia, SPUDM20 in Stockholm, the University of Toulouse, and the Wharton School of the University of Pennsylvania. This research was financed partially by a grant from the Spanish Ministerio de Educación y Ciencia.

For correspondence, please contact Robin M. Hogarth at Universitat Pompeu Fabra, Department of Economics and Business, Ramon Trias Fargas 25-27, 08005, Barcelona, Spain. (Tel: +34 93 542 2561, Fax + 34 93 542 1746). Email: robin.hogarth@upf.edu, natalia.karelaia@unil.ch

Abstract

An important problem in descriptive and prescriptive research in decision making is to identify "regions of rationality," i.e., the areas for which heuristics are and are not effective. To map the contours of such regions, we derive probabilities that heuristics identify the best of m alternatives ($m \geq 2$) characterized by k attributes or cues ($k \geq 1$). The heuristics include a single variable (lexicographic), variations of elimination-by-aspects, equal weighting, hybrids of the preceding, and models exploiting dominance. We use twenty simulated and four empirical datasets for illustration. We further provide an overview by regressing heuristic performance on factors characterizing environments. Overall, "sensible" heuristics generally yield similar choices in many environments. However, selection of the appropriate heuristic can be important in some regions (e.g., if there is low inter-correlation among attributes/cues). Since our work assumes a "hit or miss" decision criterion, we conclude by outlining extensions for exploring the effects of different loss functions.

Keywords: Decision making, Bounded rationality, Lexicographic rules, Choice theory.

JEL classification: D81, M10.

In his autobiography, Herbert Simon (1991) used the metaphor of a maze to characterize a person's life. In this metaphor, people continually face choices involving two or more alternatives, the outcomes of which cannot be perfectly predicted from the information available.[1] Extending this metaphor, the maze of choices a person faces can be thought of as a journey that crosses different regions that vary in the types of questions posed.

If endowed with unbounded rationality, one could simply calculate the optimal responses for all decisions. However, following Simon's insights, the bounded nature of human cognitive capacities implies *satisficing* mechanisms. Fortunately, this need not result in unsatisfactory outcomes if responses match the demands of the regions. However, it also raises the issue of facing the consequences of inappropriate choices.

In this paper, we characterize the maze of choices that people face as involving different "regions of rationality" where success depends on identifying decision rules that are appropriate to each region. In some regions, for example, the simplest heuristic might be sufficient (e.g., when choosing a lottery ticket). In other regions, returns to computationally demanding algorithms are potentially important (e.g., planning production in an oil refinery). What people need therefore is knowledge – or maps – that indicate the demand for rationality in different regions. In particular, since attention is the scarce resource (Simon, 1978), it is critical to know what and how much information should be sought to make decisions in different regions.

The purpose of this paper is to contribute to defining maps that characterize regions of rationality for common decisions problems. This topic is important for both descriptive and prescriptive reasons. For the former, there is a need to understand the conditions under which simple, boundedly rational decision rules – or

[1] As Simon (1991) points out, this metaphor also underlies his classic (1956) paper on what an organism needs to be able to choose effectively in given environments.

heuristics – are and are not effective. At the same time, this knowledge is critical for prescribing when people should use such rules, i.e., as decision aids. Specifically, we consider decisions between two or more alternatives based on information that is probabilistically related to the criterion of choice. The structure of these tasks can be conceptualized as involving either multiple-cue prediction or multi-attribute choice and, as such, is common. In all cases, we construct theoretical models that predict the effectiveness in different regions of several heuristics – thereby mapping the contours for which they are and are not suited.

The interest in describing implications of simple models of decision making has grown exponentially over the last five decades (see, e.g., Conlisk, 1996; Goldstein & Hogarth, 1997; Kahneman, 2003; Koehler & Harvey, 2004). Much controversy has centered on whether the heuristics that people use are effective. In particular, great interest was stimulated by research on "heuristics and biases" (Kahneman, Slovic, & Tversky, 1982) that demonstrated how simple processes can produce outcomes that deviate from normative prescriptions. Similarly, much work demonstrated that simple, statistical decision rules have greater predictive validity than unaided human judgment in a wide range of tasks (see, e.g., Dawes, Faust, & Meehl, 1989; Kleinmuntz, 1990).

An alternative view is that people possess a repertoire of boundedly rational decision rules that they apply in specific circumstances (Gigerenzer & Selten, 2001). Thus, heuristics can also produce appropriate responses. Specifically, Gigerenzer and his colleagues have demonstrated how "fast and frugal" rules can rival the predictive ability of complex algorithms (Gigerenzer, Todd, & the ABC Research Group, 1999). In their terms, such heuristics produce "ecologically rational" behavior, i.e., behavior that is appropriate in its "niche" but does not assume an underlying optimization model. What is unclear, however, is where these niches are located in the regions of rationality.

Reviewing the empirical evidence, there are clearly occasions when heuristics violate normative prescriptions as well as situations where they lead to surprisingly successful outcomes. The role of theory is to specify when both kinds of results occur or, to use the metaphor of this paper, to map the regions of rationality.

Our goal, therefore, is to develop such theory. It involves specifying analytical models for heuristics that can be used for either multi-attribute choice or multiple-cue prediction. Specifically, we derive probabilities that these models will correctly select the best of m alternatives ($m \geq 2$) based on k attributes or cues ($k \geq 1$). We also compare their effectiveness with optimizing and naïve benchmarks. The theoretical development specifies the importance of different environmental factors on heuristic performance. These include the effects of differential cue validities, inter-correlations of attributes, whether attributes/cues are measured by continuous or binary variables, levels of error in data, *and* the interactions between these factors.

This paper is organized as follows. In section I, we briefly review relevant literature. Next, in section II we specify the heuristics we examine. Section III is technical. We start by specifying the statistical theory for a heuristic based on a single continuous variable. This basic logic is then applied successively to other heuristics: first, to heuristics that make use of more than one continuous variable; and then models based on binary (0/1) variables. To facilitate presentation of key ideas, many formulas are presented in appendices. Section IV is empirical. We provide predictive tests of our models on twenty simulated and four empirical datasets. Heuristic performance is further illuminated by a regression analysis involving environmental characteristics. Our results demonstrate that no one heuristic is always best. At the same time, “sensible” heuristics yield similar results in many environments (e.g., in the presence of moderate to high inter-correlation between attributes/cues). Nonetheless, there are distinct regions where particular heuristics should be favored or avoided. The terrain, however, is complex, and our work demonstrates why theory is needed to map the regions of rationality. Finally, section V provides concluding comments as well as suggestions for further research.

I. Evidence on the predictive effectiveness of simple models

Interest in the efficacy of simple models for decision making has existed for some time with, in particular, numerous empirical demonstrations of how models based on equal (or unit) weighting schemes perform well in comparison with more complex algorithms such as multiple regression (see, e.g., Dawes & Corrigan, 1974; Dawes, 1979). Gigerenzer and Goldstein (1996) have further shown how a simple, non-compensatory lexicographic model that uses binary cues (“take the best” or TTB) is surprisingly accurate in predicting the better of two alternatives across several empirical datasets and often outperforms equal weighting (EW) (Gigerenzer, Todd et al., 1999).

Other studies have used simulation. Payne, Bettman and Johnson (1993), for example, explored tradeoffs between effort and accuracy. Using continuous variables and a weighted additive model as the criterion, they demonstrated the effects of two important environmental variables, dispersion in the weighting of variables and the extent to which choices involved dominance. (See also Thorngate, 1980). Shanteau and Thomas (2000) further defined environments as “friendly” or “unfriendly” to different models and also demonstrated similar effects through simulations.

Recently, Fasolo, McClelland, and Todd (in press) examined multi-attribute choice in a simulation using continuous variables. Their goal was to assess how well choices by models with differing numbers of attributes could match total utility and, in doing so, they varied levels of average inter-correlations among the attributes and types of weighting functions. Results showed important effects for both. With differential weighting, one attribute was sufficient to capture at least 90% of total utility. With positive inter-correlation among attributes, there was little difference between equal and differential weighting. With negative inter-correlation, however, equal weighting was sensitive to the number of attributes used (the more, the better).

Despite these demonstrations involving simulated and real data, research has lacked theoretical models for understanding how characteristics of heuristics interact with those of environments. Some work, however, has considered specific cases. Einhorn and Hogarth (1975), for example, provided a theoretical rationale for the effectiveness of equal weighting relative to multiple regression. Martignon and Hoffrage (1999; 2002) and Katsikopoulos and Martignon (in press) explored the conditions under which TTB or equal weighting should be preferred in binary choice. Hogarth and Karelaia (2004; 2005a) and Baucells, Carrasco, and Hogarth (2006) have examined why TTB and other simple heuristics perform well with binary attributes in error-free environments. And, Hogarth and Karelaia (2005b) provided a theoretical analysis for the special case of binary choice with continuous attributes.

II. Heuristics considered

Whereas fully rational models combine all relevant information in an optimal manner, heuristics use limited subsets of the same information and/or simplifying combination rules. The heuristics we examine (see Table 1) reflect these considerations and fall into three classes: (A) heuristics based on single variables or subsets of the available information; (B) equal weighting models; and (C) hybrids that combine characteristics of the two preceding classes. In addition, we consider lower and upper benchmarks: (D) heuristics that simply exploit dominance; and (E) multiple regression.

Insert Table 1 about here

We further examine how the type of data affects performance by including, where possible, versions based on both continuous and binary attributes/cues.[2] We indicate the use of the two kinds of data by suffixes: –*c* for "continuous," and –*b* for "binary".

Since most of the heuristics we examine have been considered in the literature, we limit discussion to making a few links. First, DEBA (number 3) is a deterministic version of Tversky's (1972) elimination-by-aspects (EBA) heuristic. For binary choice, this is identical to TTB of Gigerenzer and Goldstein (1996). Variables used as attributes/cues for DEBA are binary in nature and, although the amount of information consulted for each choice varies according to the characteristics of the alternatives, many decisions are based on a single attribute. In the continuous case, this is best matched by the single variable heuristic (SVc, number 1) which is equivalent to the lexicographic model investigated by Payne et al. (1993).

Second, with binary variables as cues/attributes, the EWb model predicts frequent ties. However, rather then resolving such choices at random, we use hybrid models that exploit partial knowledge. Specifically, EW/DEBA and EW/SVb are

[2] In our simulations and empirical work, we generate binary variables by median splits of the continuous variables and in this manner make direct comparisons between results based on binary and continuous variables.

models that, first, attempt to choose according to EWb. If this results in a tie, DEBA or SVb is used as a tie-breaker (see also Hogarth & Karelaia, 2005a).

Third, it is illuminating to compare performance with benchmarks. For lower or "naïve" benchmarks, we include two heuristics that simply exploit dominance, Domran (DR), numbers 8 and 9. (Simply stated, choose an alternative if it dominates the other(s). If not, choose at random.) As an upper or "normative/sophisticated" benchmark, we use multiple regression (models 10 and 11).[3]

It is important to emphasize that the heuristics differ in the demands they make on cognitive resources, specifically on prior knowledge and the amount of information to be processed. We therefore indicate, on the right of Table 1, differential requirements in terms of prior information, information to consult, calculations, and numbers of comparisons to be made (minimum to maximum). For example, Table 1 shows that the EW and DR models require no prior information other than the signs of the zero-order correlations between the cues and the criterion (a minimum). On the other hand, the lexicographic, DEBA, and hybrid models need to know which cue(s) is (are) most important. Against this, the lexicographic and DEBA heuristics do not necessarily use all cues and require no calculations. The cost of DR lies mainly in the number of comparisons that have to be made.

Our goal is to develop theoretical models that predict heuristic performance. A priori, two hypotheses can be suggested. First, we would expect models based on continuous variables to outperform their binary counterparts. Second, models that resolve ties of other models would be expected to be more accurate than the latter. Hence DEBA should be more accurate than SVb, and EW/DEBA and EW/SVb more

---

[3] We are fully aware that multiple regression is not necessarily "the" optimal model for all tasks.

accurate than EWb. However, whether DEBA is more accurate than SVc will depend on environmental characteristics.

Three types of environmental variables can be expected to affect absolute and relative model performance. These are, first, the distribution of "true" cue validities[4] (i.e., how the environment weights different variables, cf., Payne et al., 1993); second, the level of redundancy or inter-correlation among the cues; and third, the level of "noise" in the environment (i.e., its inherent predictability). Of these factors, increasing noise will undoubtedly decrease performance of all heuristics and, by extension, differences between the heuristics. Similarly, cue redundancy should decrease differences between heuristics. However, apart from these main effects, it is difficult to intuit how the different factors will combine to determine absolute and relative performance. To achieve this, we need to develop specific theory for each of the heuristics.

The next section is technical and provides the necessary theory. To facilitate exposition, we first develop the theory for one heuristic, SVc. This allows us to explain the basic theoretical logic that is subsequently applied, first, to heuristics involving more than one continuous variable, and next, to heuristics based on binary variables. In these developments, we concentrate on where the theory differs for the different heuristics. We also make use of appendices and an online supplement (Hogarth & Karelaia, 2006) to provide details of formulas and derivations.

III. Models with continuous variables

The single continuous variable heuristic (SVc). Imagine choosing from a distribution characterized by two correlated random variables, one of which is a

[4] The cue validity for a particular cue/attribute is defined by its correlation with the criterion.

criterion, Y, and the other an attribute, X. Furthermore, assume that alternative A is preferred over alternatives B and C if $y_a > y_b$ and $y_a > y_c$.[5] Now, imagine that the only information about A, B, and C are the values that they exhibit on the attribute, X. Denote these specific values by x_a, x_b, and x_c, respectively. Without loss of generality, assume that $x_a > x_b$ and $x_a > x_c$ and that the decision rule is to choose the alternative with the largest value of X, i.e., in this case A. The probability that A is in fact the correct choice can therefore be characterized by the joint probability that $Y_a > Y_b$ given that $x_a > x_b$ and $Y_a > Y_c$ conditioned on $x_a > x_c$, in other words, $P\{(Y_a > Y_b | X_a = x_a > X_b = x_b) \cap (Y_a > Y_c | X_a = x_a > X_c = x_c)\}$.

To determine this probability, assume that Y and X are both standardized normal variables, i.e., both are $N(0,1)$. Moreover, the two variables are positively correlated (if they are negatively correlated, simply multiply one by -1). Denote the correlation by the parameter ρ_{yx}, ($\rho_{yx} > 0$). Given these facts, it is possible to represent Y_i ($i=a, b, c$) by the equation:

$$Y_i = \rho_{yx} X_i + \varepsilon_i \tag{1}$$

where the ε_i ($i=a, b, c$) are normally distributed error terms, with means of 0 and variances of $(1-\rho_{yx}^2)$, independent of each other and of all X_i's.

Using equation (1), the differences between all pairs Y_i and Y_j, $j\neq i$, can be written as

$$Y_i - Y_j = \rho_{yx}(X_i - X_j) + (\varepsilon_i - \varepsilon_j). \tag{2}$$

Thus, $Y_i > Y_j$ for all $j\neq i$ if

$$\varepsilon_j - \varepsilon_i < \rho_{yx}(X_i - X_j) \qquad \text{for all } j\neq i. \tag{3}$$

[5] We denote random variables by upper case letters, e.g., Y and X, and specific values or realizations by lower case letters, e.g., y and x. As an exception, we use lower case Greek letters to denote random error variables, e.g., ε.

The probability of correct choice can now be restated as

$$P\{(\varepsilon_b - \varepsilon_a < \rho_{yx}(x_a - x_b)) \cap (\varepsilon_c - \varepsilon_a < \rho_{yx}(x_a - x_c))\}. \tag{4}$$

To determine this probability, we make use of the facts that the differences between the error terms, (ε_b - ε_a) and (ε_c - ε_a), are both normally distributed with means of 0 and variances of $2(1-\rho_{yx}^2)$. Standardizing (ε_b - ε_a) and (ε_c - ε_a), we can re-express equation (4) as

$$P\{(z_1 < l_{ab}) \cap (z_2 < l_{ac})\}, \tag{5}$$

where z_1 and z_2 are standardized bivariate normal variables with variance / covariance $V_z = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$, and means of 0, and $l_{ab} = \frac{\rho_{yx}(x_a - x_b)}{\sqrt{2(1-\rho_{yx}^2)}}$, $l_{ac} = \frac{\rho_{yx}(x_a - x_c)}{\sqrt{2(1-\rho_{yx}^2)}}$.[6]

Therefore, the probability of correctly selecting A over B and C is

$$\int_{-\infty}^{l_{ab}} \int_{-\infty}^{l_{ac}} \varphi(z|\mu_z, V_z) dz_1 dz_2 \tag{6}$$

where the probability density function (pdf) $\varphi(z|\mu_z, V_z) = \frac{|V_z|^{1/2}}{(2\pi)^{(m-1)/2}} e^{-\frac{1}{2}z' V_z^{-1} z}$, $z' = (z_1, z_2)$, with m (the number of alternatives) equal to 3.

As can be seen by observing the limits l_{ab} and l_{ac}, the probability of correct choice increases when the correlation ρ_{yx} becomes larger as well as when the differences $(x_a - x_b)$ and $(x_a - x_c)$ are greater.[7]

To generalize the above, assume that there are m (m > 3) alternatives from which to choose and that each has a specific X value, x_i, i=1,…., m. Without loss of generality, assume that x_1 has the largest value and we wish to know the probability

[6] The derivation that $\sigma_{z_1,z_2} = 1/2$ can be found in the online supplement (Hogarth & Karelaia, 2006).

[7] This is illustrated graphically in the online supplement (Hogarth & Karelaia, 2006).

that the corresponding alternative has the largest value on the criterion. Generalizing from the above, this probability can be calculated using properties of the multivariate normal distribution and written,

$$\int_{-\infty}^{d_1^*} \cdots \int_{-\infty}^{d_{m-1}^*} \varphi(z|\mu_z, V_z) dz_1 \dots dz_{m-1} \tag{7}$$

where the pdf $\varphi(z|\mu_z, V_z)$ is defined as above; $d_t^* = \dfrac{\rho_{yx} d_t}{\sqrt{2(1-\rho_{yx}^2)}}$ for each between-alternative comparison $t = \overline{1, m-1}$; the elements of $z' = (z_1, z_2, ..., z_{m-1})$ are jointly distributed standard normal variables, and V_z^{-1} is the inverse of the (*m-1*)x(*m-1*) variance-covariance matrix where each diagonal element is equal to 1 and all off-diagonal elements equal ½. For binary choice, that is, when *m* = 2, analogous derivations lead to similar expressions to those shown above (see Hogarth & Karelaia, 2005b).

The probabilities given above are those associated with particular observations, i.e., that A is larger than B and C given that a specific value, x_a, exceeds specific values x_b and x_c. However, it is also instructive to consider the overall expected accuracy of SV, i.e., the overall probability that SV makes the correct choice when sampling at random from the population of alternatives.

Overall, SV can make successful choices in three ways: selecting A when x_a is bigger than x_b and x_c; selecting B when x_b is bigger than x_a and x_c; and selecting C when x_c is bigger than x_a and x_b. Since the three cases are symmetric, the overall probability of correct choice is $3P\{((X_a > X_b) \cap (X_a > X_c)) \cap ((Y_a > Y_b) \cap (Y_a > Y_c))\}$.

To derive analytically the overall probability of correct choice by SV when sampling at random from the underlying population of alternatives, the latter expression should be integrated across all possible values that can be taken by $D_1 = X_a - X_b > 0$, and $D_2 = X_a - X_c > 0$. That is

$$3\int_0^{\infty}\int_0^{\infty} \varphi(d|\mu_d, V_d)\left[\int_{-\infty}^{d_1^*}\int_{-\infty}^{d_2^*} \varphi(z|\mu_z, V_z)dz_1 dz_2\right]dd_1 dd_2 \tag{8}$$

where pdfs $\varphi(d|\mu_d, V_d)$ and $\varphi(z|\mu_z, V_z)$ are defined for $z' = (z_1, z_2)$, $V_z = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$,

$d' = (d_1, d_2)$, $V_d = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, and $d_t^* = \dfrac{\rho_{yx} d_t}{\sqrt{2(1-\rho_{yx}^2)}}$ for $t = \overline{1,2}$. In Appendix A, we generalize these formulas for choosing one of m alternatives.

Equal weighting (EWc) and multiple regression (MRc). What are the predictive accuracies of models that make use of several, k, continuous cues or variables, $k > 1$? We consider two models. One is equal weighting (EWc). The other is multiple regression (MRc). To analyze these models, assume that the criterion variable, Y, can be expressed as a function of k predictor variables, X_j ($j = 1,..,k$) each of which is $N(0,1)$. For EWc, the predicted Y value associated with any vector of observed x's is equal to $\frac{1}{k}\sum_{j=1}^{k} x_j$ or $\bar{x}$. Similarly, the analogous prediction in MRc is given by $\sum_{j=1}^{k} b_j x_j$ or $\hat{y}$ where the b_j's are estimated regression coefficients. In using these models, therefore, the decision rules are to choose according to the largest $\bar{x}$ for EWc and the largest $\hat{y}$ value for MRc.

How likely are EWc and MRc to make the correct choice? The rationale to derive the probabilities of correct choice is the same as in the single variable (SVc) case. All formulas are shown in Appendix A (for choosing the best of m alternatives). In particular, we present there the elements that are specific to different models, such as the initial formulations (corresponding to equation 1), the error variances, the upper limits of integration d_t^* used to calculate probabilities when applying the analogues to equations (6) and (8), and finally the variance-covariance matrix V_d .

Models with binary variables. To discuss expected predictive performance of heuristics based on binary variables, assume that the dependent variable, Y, is generated by a linear model of the form

$$Y = a + \sum_{j=1}^{k} \gamma_j W_j + \varpi \tag{9}$$

where $W_j = 0, 1$ are the binary variables ($j = 1,\dots,k$), the γ_j are weighting parameters and ϖ is a normally distributed error term.

To derive theoretical predictions, we adopt an approach similar to that used with continuous variables. We therefore focus on issues that differ between the continuous and binary cases.

The single binary variable heuristic (SVb). Assuming that $w_a > w_b$ and $w_a > w_c$, the probability that SVb chooses correctly between three alternatives, A, B, and C is $P\{(Y_a > Y_b | W_a = w_a > W_b = w_b) \cap (Y_a > Y_c | W_a = w_a > W_c = w_c)\}$.

To determine this probability, recall that Y is a standardized normal variable $N(0,1)$. The binary variable, W, however, only takes values of 0 and 1 and thus has a mean of $\frac{1}{2}$ and standard deviation, σ_w , of $\frac{1}{2}$.[8] Denoting the correlation between Y and W by ρ_{yw}, we can express Y as being equal to $a_{SVb} + \frac{\rho_{yw}}{\sigma_w} W + \varsigma$, or, simply,

$$Y = a_{SVb} + 2\rho_{yw} W + \varsigma \tag{10}$$

where ς is a normally distributed error term $N(0,\ 1 - \rho_{yw}^2)$.[9]

[8] Recall that binary variables are created by median splits of continuous variables.

[9] Since E(Y) = 0, it follows that the intercept $a_{SVb} = -\rho_{yw}$.

Proceeding in similar fashion to the continuous case, we find that the probability of SVb predicting correctly is given by equation (6) but with different upper integration limits: $l_{ab} = l_{ac} = \frac{2\rho_{yw}}{\sqrt{2(1-\rho_{yw}^2)}}$. (The two limits are the same since both $(w_a - w_b)$ and $(w_a - w_c)$ are equal to one).

Since the only difference between the theoretical expressions for the continuous and binary cases lies in the formulas for the upper limits of integration, generalizing the above for choices between *m* (*m* > 3) alternatives is analogous to that for the continuous case.

Following the same rationale, we can derive the formulas for the probabilities for EWb and MRb when choosing the best of three alternatives using binary variables. In Appendix B, we present the key formulas for the models using binary variables, analogous to those in Appendix A for the models using continuous variables.

The DEBA heuristic. Recall that this multi-stage model uses binary cues and works in the following way. At the first stage, alternatives with values of 0 for the most important cue are eliminated unless all alternatives exhibit 0. If only one alternative has a value of 1, it is selected and the process terminates. If, however, more than one alternative remains, the same procedure takes place with the remaining alternatives except that the second most important cue is used. The process continues in the same manner through subsequent stages, if necessary. It stops when either only one alternative remains (i.e., the chosen alternative) or, if there is more than one alternative but no more cues, choice is determined at random among the remaining alternatives.

The probability that a given alternative was chosen correctly by DEBA is the probability that the sequence of decisions (or eliminations) made by the model at each stage is correct. Thus, since at each stage of the model decisions are made conditional on the preceding stages, the key parameters in estimating these probabilities are the partial correlations between *Y* and W_j, $j = 1,\ldots,k$ (i.e., controlling for previous stages). For the first stage, this is ρ_{yw_1}, for the second $\rho_{yw_2.w_1}$, for the third, $\rho_{yw_3.w_1w_2}$, and so on.[10] (Two examples of the calculations for DEBA are provided in Appendix C.)

Overall, the probability of DEBA making the correct choice has to be calculated on a case-by-case basis taking into account, at each stage, the probability that the selected alternative should be chosen over the alternative(s) eliminated at that stage using the partial correlation of the cue appropriate to the stage. Moreover, the probability for each case includes the probabilities of successful decisions at each stage. If at the final stage, there are two or more alternatives, the appropriate random probability is adjusted by the probability that correct decisions were taken at previous stages (see the examples in Appendix C).[11]

The EW-SVb heuristic. The first stage of this heuristic uses EWb. If a single alternative is chosen, the probability of it being correct is found by applying the formula for EWb. If two or more alternatives are tied, a second stage consists of selecting the alternative favored by the first cue. To calculate the probability that this is correct, one needs to calculate the joint probability that the selected alternative (a) is larger than the alternatives eliminated at the first stage, and (b) larger than the other alternatives considered at the second stage. (To calculate these probabilities, use is made of the appropriate analogues to equation 6). Any ties remaining after stage two

---

[10] For example, $\rho_{yw_2.w_1} = \dfrac{\rho_{yw_2} - \rho_{yw_1}\rho_{w_1w_2}}{\sqrt{\left(1-\rho_{yw_1}^2\right)\left(1-\rho_{w_1w_2}^2\right)}}$.

[11] In this section, we have only indicated the general strategy for calculating relevant probabilities for DEBA. The details involve repeated applications of the same probability theory principles applied in many different situations (Karelaia & Hogarth, in preparation).

are resolved at random with a corresponding adjustment being made to the probability calculations.

The EW-DEBA heuristic. This starts as EWb. If EWb chooses one alternative, the probability of correct choice of the model coincides with that for EWb. If two or more alternatives are tied, the DEBA model is used to choose between the remaining alternatives and probabilities are calculated accordingly (see above).

IV. Empirical evidence

Our equations provide exact theoretical probabilities for assessing the performance of the different heuristics in specified conditions, i.e., to map the contours of the regions of rationality. However, several factors affect absolute and relative performance levels of the models (e.g., cue validities, inter-correlation among variables, continuous vs. binary variables, error), and it is difficult to assess their importance simply by inspecting the formulas.

We therefore use both simulated and empirical data to illuminate model performance under different conditions. Real data have the advantage of testing the theory in specific, albeit limited environments. Simulated data, on the other hand, facilitate testing model predictions over a wide range of environments. We first consider the simulated data.

Simulation design and method. The simulation design used for choosing the best from two, three and four alternatives is presented in the upper panel of Table 2. Our goal was to vary environmental factors that we thought, a priori, would be important. These were the level of noise or error in the environment, redundancy between attributes/cues, the distribution of cue validities, and the number of cues. A priori we would have liked to vary these factors orthogonally. However, correlations between factors restrict implementing a fully systematic design.

Overall, we specified 20 different populations that are subdivided into four sets or cases – A, B, C, and D – each of which contains five sub-cases (labeled 1, 2, 3, 4, and 5). We therefore varied some factors at the level of the cases (A, B, C, and D) and others across sub-cases (i.e., within A, B, C, and D).

At the level of cases, A and B involved three cues or attributes whereas cases C and D involved five. Cases A and C had little or no inter-cue correlation; cases B and D had moderate to high intercorrelation.

Insert Table 2 about here

Across sub-cases (i.e., from 1 through 5 within each of A, B, C, and D), we varied: (1) the variability of cue validities (maximum less minimum); (2) the validity of the first (i.e., most important) cue; (3) average validity; and (4) the correlation between y and $\bar{x}$. For all, values increase from the sub-cases 1 through 5. As a consequence, the R^2 on initial fit for MR also increases across sub-cases. This implies that the sub-cases 1 involve high levels of error whereas the sub-cases 5 are, in principle, quite predictable environments. Sub-cases 2, 3, and 4 fall between these extremes.

To conduct the simulation, we defined 20 sets of standardized multivariate normal distributions with the parameters specified in Table 2 and generated samples of size 40 from each of these populations. The observations in each sample were split at random on a 50/50 basis into fitting and prediction sub-samples and model parameters were estimated on the fitting sub-sample.[12] Two, three or four alternatives (as appropriate) were then drawn at random from this sub-sample and, using the estimated model parameters, probabilities of correctly selecting the best of these specific alternatives were calculated. This was then compared to what actually happened, that is, on a fitting basis. Next, alternatives were drawn at random from the prediction sub-sample, relevant probabilities calculated using the parameters from the fitting sub-sample, and predictions compared to realizations. This exercise was repeated 5,000 times (for each of the choices involving two, three, and four alternatives).[13]

The above describes the procedure used for continuous data. For models using binary data, we followed exactly the same procedures except that predictor variables only took values of 0 or 1. Specifically, since we were sampling continuous normalized variables, we created binary variables by median splits (i.e., binary variables were set to 0 for negative values of continuous variables and 1 for non-negative values). Thus, if one estimates the parameters of the 20 populations for the binary data, the estimates differ systematically from their continuous counterparts.

Simulation results. First of all, the models were successful on cross-validation: the theoretical predictions (i.e., formula-based) matched the actual realizations (i.e., simulated) almost perfectly. The only exception was MR which manifested the well-known over-fitting effect. Although we used adjusted R^2 in

[12] It is of particular interest to examine the performance of heuristics in situations involving small samples. We therefore chose sample sizes of 40 to allow for fitting and holdout samples of size 20.
[13] A possible criticism of our predictive tests of the single variable models (SVc, SVb, and DEBA) is that we did not use the sampling process to determine the most important variable (for SVc and SVb) nor the rank orders of the cue validities (for DEBA). Instead, we endowed the models with the appropriate knowledge. However, in subsequent simulations we have found that with sample sizes of 20 (as here) the net effect of failing to identify the most important variable is quite small. Similarly, as long as DEBA correctly identifies the most important variable, net differences are also small (Hogarth & Karelaia, 2005a).

making predictions, the adjustment was insufficient.[14] We therefore report here results based only on actual realizations in the holdout samples. These are presented in the lower panel of Table 2 for the choice of the best of three alternatives. The analogous results for choosing from two and four alternatives can be found in the online supplement (Hogarth & Karelaia, 2006).[15] Qualitatively, the relative effectiveness of the models is similar whether one considers the best of three (Table 2) or two or four (Hogarth & Karelaia, 2006). What changes, of course, is the general level of performance which diminishes as the number of alternatives increases.

To simplify reading the table, note that the best level of performance for each population (i.e., per column) is presented in bold (e.g., **52**). In addition, when MRc is best, we also denote the second best in bold. We further show means for each column and row. The column means thus represent the average results of all models within specific populations whereas the row means characterize average model performance across populations.

There are several systematic effects in the results presented in Table 2. First, across all environments, the mean levels of heuristic performance do not vary greatly, i.e., on the right-hand side column of Table 2 the means (excluding MRc) range only between 43% and 57%.

Second, consider whether predictor variables are binary or continuous. The use of binary as opposed to continuous variables implies a loss of information. As such, we expected that models based on continuous variables would predict better than their binary counterparts. Indeed, this is always the case in three direct comparisons: SVc vs. SVb, EWc vs. EWb, and MRc vs. MRb. Specifically, note that SVb and EWb are

[14] One can also argue with some justification that the ratio of observations to predictor variables is too small to use multiple regression (particularly for cases C and D). However, we are particularly interested in observing how well the different models work in environments where there are not many observations.

[15] This also provides details of how well all models both fit and predicted the different cases.

both handicapped relative to SVc and EWc in that they necessarily predict many ties that are resolved at random. Thus, so long as the knowledge in SVc and EWc implies better than random predictions, models based on continuous variables are favored. On the other hand, the performance of DRb dominates DRc. We comment on this result below.

Third, models that resolve ties generally perform better than their counterparts that are unable to do so (e.g., DEBA vs. SVb, and EW/DEBA and EW/SVb vs. EW). However, in the presence of redundancy, these differences are quite small and can even reverse (i.e., see cases B and D).

Fourth, an interesting comparison is that between SVc and DEBA. The former uses a single, continuous variable. The latter relies heavily on one binary variable but can also use others depending on circumstances. It is thus not clear which strategy actually uses more information. However, once again, characteristics of the environment determine which strategy is more successful. SVc dominates DEBA in case B as well as for much of cases A and D. On the other hand, DEBA dominates SVc in case C.

Insert Figure 1 about here

Fifth, Figure 1 illustrates significant differences in some regions by depicting the performance of five heuristics across all the environments for the choice of the best of three (i.e., data from Table 2). Two of these models, DRb and MRc, represent, respectively, naïve and "sophisticated" benchmarks. The others, SVc, DEBA, and EWc are different types of heuristics (see Table 1). Both SVc and DEBA require prior knowledge of what is important (DEBA more so than SVc). However, they use little information and neither involves any computation. (DEBA, it should be recalled, also operates on binary data.) EWc, on the other hand, does not require knowledge of differential importance of variables but does use all information available and needs some computational ability.

In interpreting Figure 1, it is instructive to recall that cases A and C (on the left) represent environments with low redundancy whereas cases B and D (on the right) have higher levels of redundancy. Also within each case, the amount of noise in the environment decreases as one moves from sub-case 1 (on the left) to sub-case 5 (on the right).

As expected, in the noisier environments (sub-cases 1), the performances of all models are degraded such that differences are small. However, as error decreases (i.e., moving right toward sub-cases 5), model performances vary by environmental conditions. With low redundancy (cases A and C), there appear to be large differences in model performance. However, in the presence of redundancy (cases B and D), there are two distinct classes of models: SVc and MRc have similar performance levels and are superior to the others. We further note that SVc is most effective in case B and also does well as environmental predictability increases in cases A and D. DEBA is never the best model but performs quite adequately in case C where EWc has the best performance. Of the benchmark models, DRb generally lags behind the other models (as would be expected). Finally, although MRc is typically one of the better models, it does not dominate in all environments.

Discussion of simulation. We consider three issues: first, the fact that the overall performance of the heuristics was more similar than might have been expected a priori; second, the intriguing result that the DOMRAN heuristic performed better using binary as opposed to continuous variables; and third, factors that explain the differential performance of the heuristics.

One way of examining similarity in performance between heuristics is to examine the percentage of cases for which they make the same predictions (whether correct or incorrect). Limiting ourselves only to those heuristics illustrated in Figure 2, we find that the average across all pairs of heuristics and datasets is 63% – in other words, these heuristics made the same predictions in some two-thirds of cases.[16] This average is, of course, higher when cues/attributes are more inter-correlated (means of 66% for cases B and D but 62% for A and 57% for C – data not shown in tables). At one level, this result might seem surprising. On the other hand, our heuristics are "sensible" in the sense that they all use *valid* variables. Where they differ is in how many they use and how these are weighted.

Second, the surprising result that DRb outperforms DRc can be resolved by noting that the former exploits more cases of "apparent" dominance than the latter. Specifically, in choosing the best of three, DRb was only forced to make random choices in, on average, 51% of all cases whereas the same figure was 81% for DRc (these data are not shown in our tables). Thus, to the extent that the "additional" dominance cases detected by DRb had more than a random chance of being correct, DRb outpredicted DRc. This finding is important because it demonstrates how a simple strategy can exploit the structure of the environment such that more information (in the form of continuous as opposed to binary variables) does not improve performance (a so-called "less is more" effect, Goldstein & Gigerenzer, 2002; Hertwig & Todd, 2003).

Third, a way of summarizing factors that affect the performance of the different models is to consider the regression of model performance on statistics that

[16] To calibrate this result, note that if all models chose at random, the probability of any two models agreeing when selecting one of three alternatives would be 1/3.

characterize the 20 simulated environments. In other words, consider regressions for each of our eleven models of the form

$$P_i = Z\underline{\delta_i} + \tau_i \quad (11)$$

where P_i is the performance realization of model *i* (*i* = 1, …, 11); *Z* is the (20 x 3) x *s* matrix of independent variables (statistics characterizing the datasets where there are three choice situations, i.e., best of two, three, or four alternatives); $\underline{\delta_i}$ is the *s* x 1 vector of regression coefficients; and τ_i is a normally distributed error term with constant variance, independent of *Z*.

Insert Table 3 about here

To characterize the environments or datasets (the *Z* matrices), we chose the following variables: variability of cue validities (maximum less minimum), the validity of the most important cue (r_{yx_1}), the validity of the average of the cues ($r_{y\bar{x}}$), average inter-correlation of the cues, average validity of the cues, number of cues, and R^2 for MRc and MRb.[17] We also used dummy variables to model the effects of choosing between different numbers of alternatives. Dummy1 captures the effect of choosing from three as opposed to two alternatives, and Dummy2 the additional effect of choosing from four alternatives. Results of the regression analyses are summarized in Table 3.

We used a step-wise procedure with entry (exit) thresholds for the variables of <.05 (>.10) for the probability of the F statistic. All coefficients for the models shown

[17] We only used R^2 as an independent variable for MRc and MRb because we thought it would be appropriate for these models. For the other models, however, it was deemed more illuminating to characterize performance by the other measures (R^2 and these other measures are correlated in different ways). It should also be noted that we used the same statistics (based on continuous variables) to characterize the environments for models using both continuous and binary variables on the grounds that the underlying environments were based on continuous variables.

in Table 3 are statistically significant (most with $p < .001$) and all regressions fit the data well (see R^2 and estimated standard errors at the foot of Table 3). The constant terms measure the level of performance expected of the models in binary choice absent information about the environment (approximately 45, i.e., from 39 to 50). Dummy1 indicates how much performance would fall when choosing between three alternatives (between 11 and 15), and Dummy2 shows the additional drop experienced when choosing among four alternatives (between 6 and 9).

For SVc and SVb, only one other variable is significant, the correlation between the single variable and the criterion. This makes intuitive sense as does the fact that the regression coefficient is larger with continuous as opposed to binary variables (50 vs. 29). The DEBA and EW heuristics are all heavily influenced by the correlation between the criterion and $\bar{x}$. Recall, however, that this correlation is itself an increasing function of average cue validity and the number of cues but decreasing in the inter-correlation between cues (see footnote 1 in Appendix A). Thus, *ceteris paribus*, increasing inter-correlation between the cues reduces the absolute performance levels of these heuristics. DEBA differs from the EW heuristics in that the correlation of the most valid cue is a significant predictor. This matches expectations in that DEBA relies heavily on the validity of the most important cue whereas EW weights all cues equally. (We also note that the SV heuristics weight the most valid cue more heavily than DEBA.) As to Domran, the interpretation of the signs of all coefficients is not obvious. Finally, for MRc it comes as little surprise that R^2 should be so important although this is less salient for MRb.

A possible surprise is that variability in cue validities (maximum less minimum) was not a significant factor for most models. One might have thought, a priori, that such dispersion would have been important for DEBA (cf., Payne et al., 1993). This is not the case; but it is consistent with theoretical analyses of DEBA that demonstrate its robustness relative to different "weighting functions" (Hogarth & Karelaia, 2005a; Baucells, Carrasco, & Hogarth, 2006).

Finally, whereas the regression statistics paint an interesting picture of model performance in the particular environments observed, we caution against overgeneralization. We only observed restricted ranges of the environmental statistics (i.e., characteristics) and thus cannot comment on what might happen beyond these ranges. Our approach, however, does suggest how to illuminate model x environment interactions.

To summarize, across all 20 environments that, inter alia, are subject to different levels of error, the relative performances of the different models were not seen to vary greatly when faced with the same tasks (e.g., choose best of three alternatives). However, there were systematic differences due to interactions between characteristics of models and environments. Thus, whereas the additional information contained in continuous as opposed to binary variables benefits some models, e.g., SV and EW, it can be detrimental to others, e.g., DR. Second, models varied in the extent to which they were affected by specific environmental characteristics. SV models, for example, depend heavily on the validity of the most valid cue whereas this only affects EW models through its impact on average cue validity. Interestingly, the validity of the average of the cues was seen to have more impact on the performance of DEBA than the validity of the most valid cue. Average inter-correlation of predictors or redundancy tends to reduce performance of all models (except SV).

Overall, results do match some general trends noted in previous simulations (Payne et al., 1993; Fasolo et al., in press); however, patterns are not simple to describe. The value of our work, therefore, is that we now possess the means to make precise theoretical predictions for various heuristics in different environments. That is, given specified environments, we can predict absolute and relative model performance *a priori.*

Empirical data. We used datasets from three different areas of activity. The first involved performance data of the 60 leading golfers in 2003 classified by the Professional Golf Association (PGA) in the USA.[18] From these data (N = 60), we examined two dependent variables: "all-round ranking" and "total earnings." The first is a measure based on eight performance statistics. For our models, we chose three predictor variables that account for 67% of the variance in the criterion. These were mean numbers – across rounds played – of birdies, total scores, and putts. Since the first variable was negatively related to the criterion, it was rescaled (multiplying by -1).

Eighty-two percent of the variance in the second golf criterion, total earnings, could be explained by three variables, "number of top 10 finishes," "all-round ranking" (the previous dependent variable), and "number of consecutive cuts." Of these, all-round ranking was negatively correlated with the criterion and so rescaled (multiplying by -1).

The second dataset consisted of rankings of PhD economics programs in the USA on the basis of a 1993 study by the National Research Council (N = 107).[19] Three variables accounted for 80% of the variance in the rankings: number of PhD's produced by programs for the academic years 1987-88 through 1991-1992, total number of program citations in the period 1988-1992 divided by number of program faculty, and percentage of faculty with research support.

---

[18] These data were obtained from the webpage http://www.pgatour.com/stats/leaders/r/2003/120 (accessed in June 2004). They are performance statistics of golfers in the main PGA Tour for 2003.
[19] For more details, see the webpage http://www.phds.org (accessed in June 2004).

The third dataset was taken from the UK consumer organization *Which?*'s assessments of digital cameras in 2004 (N = 49).[20] Three variables were found to explain 72% of the variance in total test scores: image quality, picture downloading time, and focusing.

Insert Table 4 and Figure 2 about here

Table 4 (upper panel) summarizes statistical characteristics of these datasets. As can be seen, there is little to moderate variability in cue validities (compare *Economics PhD programs* with the other datasets); all datasets have at least one highly valid cue; average inter-correlation varies from moderate to low (*Consumer reports: Digital cameras*); and there are high correlations between the criteria and the means of the predictor variables.

The testing procedure was similar to the simulation methodology. We divided each sample at random on a 50/50 basis into fitting and testing sub-samples. Model parameters were then estimated on the fitting sub-sample and these parameters used to calculate the probabilities that the models would correctly choose the best of two, three, and four alternatives that had also been randomly drawn from the same sub-sample. This was the fitting exercise. To test the models' predictive abilities, two, three, and four alternatives were drawn at random from the second or testing sub-sample and, using the parameters estimated from the data in the first or fitting sub-sample, model probabilities were calculated for the specific cases and subsequently compared to realizations. This exercise was repeated 5,000 times such that the data reported in Table 4 (lower panel) shows the aggregation of all these cases. Also

[20] For further details, see the webpage http://sub.which.net (accessed in June 2004).

similar to the simulation methodology, we created binary datasets by using median splits of the continuous variables.

Once again, there is an excellent fit between predictions and realizations (except for MR). For this reason, in Table 4 (lower panel), we only report the results of realizations for choosing the best out of two, three, and four alternatives for each dataset.[21] Overall, the models have high levels of performance which, naturally, diminish as the number of alternatives increases (however, perhaps, not as much as might have been thought *a priori*. See, in particular, *Economics PhD programs*).

As with the simulated datasets, there is much agreement between the predictions of the different heuristics.[22] Moreover, the SV, EW, and MR models with continuous variables outperform their binary counterparts (with one exception in 36 comparisons) but DRb dominates DRc. (Once again, DRb identifies "dominance" more often than DRc, 59% versus 45% for choosing the best of three). As to the SVc versus DEBA comparison, SVc outperforms DEBA by some margin on the *Golf rankings* and *Golf earnings* datasets but DEBA dominates SVc on the other two datasets. From a statistical viewpoint, these two datasets differ from the others in that the *Economics PhD programs* has low variability of cue validities and *Consumer reports: Digital Cameras* has low average cue inter-correlation.

Figure 2 shows the performances of SVc, DEBA, EWc, DRb, and MRc. Whereas the datasets have some similarities (e.g., the validities of the first cue and the correlations between the criteria and the means of the predictor variables are almost the same), other differences (notably variability in cue validities and levels of cue inter-correlations) are sufficient to change relative predictive performance. SVc is

[21] Details of model fits and predictions are included in the online supplement (Hogarth & Karelaia, 2006).

[22] For example, for choosing the best of three, the average of the pairwise agreements between the heuristics in Figure 2 is 74% (data not shown here).

effective for *Golf rankings* and *Golf earnings*, EWc predicts the *Economics PhD programs* well and DEBA is best for *Consumer reports: Digital cameras*. As to the benchmark models, DRb generally has the lowest performance, with the exception of choosing one of two in the *Consumer reports: Digital cameras* and *Economics PhD programs* datasets where it outperforms SVc. For all datasets, MRc is always one of the better models but it does not dominate the other models. Taken as a whole, our theoretical models account for complex patterns of data.

V. Discussion

We have mapped regions of rationality by studying a class of decisions that involve choosing the best of m ($m \geq 2$) alternatives on the basis of k ($k \geq 1$) cues or attributes. As such, this is a common task in inference and also has applications to preference (cf., Hogarth & Karelaia, 2005a). We have shown – through theory, simulation, and empirical demonstration – that certain simple, heuristics can have effective performance relative to more complex, sophisticated benchmarks and, indeed, when data are scarce can, on occasion, perform better than the latter. More importantly, our theoretical analysis predicted differential model performance in a wide range of environments. Thus, for example, for our empirical datasets we predicted – and later verified – that EWc would be the best of the simple models for *Economics PhD programs* but SVc the best for *Golf rankings*.

General trends concerning relative heuristic performance have, of course, been known for some time (e.g., effects of inter-correlations between cues or attributes). However, the advantage of our approach is that we can specify *a priori* the *combined* effects of different environmental characteristics such as variability in cue validities, inter-correlations, level of error, and so on. Moreover, we observed that the effects of tradeoffs between such factors are complex and often defy simple description. The terrain that we have mapped has many dimensions.

One factor we did not consider was the effect of sampling alternatives from the underlying populations in biased or non-random ways. Clearly, results would be different if sampling excluded certain profiles of alternatives such as those likely to dominate others or be dominated.[23] On the other hand, our theoretical method allows us to make case-by-case predictions such that – through suitable aggregation – we could make predictions for samples drawn in specific, non-random ways provided the same sampling procedures are used in both fitting and holdout samples. Showing the effects of such non-random sampling is thus straightforward and can be addressed in future research.

This paper is also limited by the criterion used to measure model effectiveness, i.e., the emphasis on probability of correct choices. This might seem restrictive in that it assumes a "hit or miss" criterion with no consideration as to how "good" the other alternatives are. We accept this limitation but emphasize that our methodology can be easily extended to other loss functions.

First, for simplicity, consider an example of binary choice using a single variable (SV) where our methodology is used to determine the probability that alternative A is better than alternative B. In addition to determining the probability that Y_a is greater than Y_b (or equivalently $Y_a - Y_b > 0$), we can also consider the probability that $Y_a - Y_b > c$ where $c > 0$. To find this value, we need to modify the inequalities involving error differences. For example, the inequality (3), that for A and B is written as $\varepsilon_b - \varepsilon_a < \rho_{yx}\left(X_a - X_b\right)$, becomes:

[23] Note that we would not be able to apply our "overall formulas" (e.g., equation 8) to these populations because they assume unbiased, random sampling.

$$\varepsilon_b - \varepsilon_a < \rho_{yx}\left(X_a - X_b\right) - c \qquad (3')$$

and we proceed as before with the calculations. Moreover, by repeating this calculation for different values of *c*, one can investigate *how much* better Y_a is likely to be compared to Y_b. As an example, one can calculate the value of *c* for which $P\{Y_a - Y_b > c\} > 0.5$ or other meaningful levels of probability.

Second, our theoretical models can be used to specify not just the probability that one alternative will be correctly selected but also the probabilities for all alternatives. For example, imagine choosing between three alternatives A, B, and C using the SV model and having observed $x_a > x_b > x_c$. Above, we calculated the probability $P\{(Y_a > Y_b | X_a = x_a > X_b = x_b) \cap (Y_a > Y_c | X_a = x_a > X_c = x_c)\}$. However, we could also have calculated the probability that B is the largest, that is $P\{(Y_b > Y_a | X_a = x_a > X_b = x_b) \cap (Y_b > Y_c | X_b = x_b > X_c = x_c)\}$ and so on. In other words, we can specify the probabilities associated with all possibilities. Given such distributions over possible outcomes, it is straightforward to consider the effects of different loss functions, a topic we also leave for further research.

Future work could also build on our theoretical approach to consider variations of the models we have examined here. One might analyze, for example, models that are less "frugal" than SV in that they use more than one cue and yet still do not use all available information (e.g., Lee & Cummins, 2004; Karelaia, in press). As another example, models might involve mixtures of categorical and continuous variables or the effects of different types of error. How, for instance, would heuristics perform when there are errors in the variables (perhaps due to measurement problems) or missing values? In addition, it will be important to investigate effects due to deviations from assumptions of normal distributions examined in this paper. Clearly many further elaborations can be undertaken.

Our work has particular implications for decision making when attention is a scarce resource. As stated by Simon (1978):

> In a world in which information is relatively scarce, and where problems for decision are few and simple, information is almost always a positive good. In a world where attention is a major scarce resource, information may be an expensive luxury, for it may turn our attention from what is important to what is unimportant. We cannot afford to attend to information simply because it is there (Simon, 1978, p. 13).

By way of illustration, Simon described executives whose management information systems provide excessive, detailed information. Our work also identified regions where more information does not necessarily lead to better decisions and, if we assume that more complex models require more cognitive effort (or computational cost), there are many areas where there is no tradeoff between accuracy and effort. For example, in cases B and D illustrated in Figure 2, the simple SVc model is more accurate than the other models indicated across almost the whole range of conditions and, yet, it uses less information. On the other hand, EWc is generally best in case C where SVc lags behind the other models. However, EWc uses more information than both SVc and DEBA such that one can ask whether the additional predictive ability is worth its cost.

The models we examined might also be used in applied areas such as consumer research (cf., Bettman, Luce, & Payne, 1998). That is, instead of assuming that consumers make tradeoffs across many attributes, simpler SV or EW models can be constructed after eliciting a few simple questions concerning, say, relative importance of attributes. For example, we have shown elsewhere that if people have loose preferences characterized by binary attributes, the outputs of DEBA are remarkably consistent with more complex, linear tradeoff models (Hogarth &

Karelaia, 2005a). However, it would be a mistake to assume that consumer preferences can always be modeled by one simple model (e.g., EW). Indeed, our theoretical analysis provides the basis for deciding which models are suited to different environments.

As suggested, our models clearly have many implications for prescriptive work. In addition to determining when heuristic models are appropriate for specific, applied problems in, for example, forecasting, personnel assessment and recruitment, medical decision making, etc, heuristics are useful supports for more complicated, decision analytical modeling. When, for instance, does one need to assess tradeoffs precisely, or can heuristic-based simplifications suffice? Here theory such as developed in this paper has great practical use for determining which heuristic to use, and when.

Finally, in a world where attention is the scarce resource, we note that "rational behavior" consists of finding the appropriate match between a decision rule and the task with which it is confronted – a principle that is valid for both descriptive and prescriptive approaches to decision making. On considering both dimensions, therefore, we do not need to assume unlimited computational capacity. However, by relaxing this assumption, we incur two costs. The first, analyzed in this paper, is to identify the task conditions under which specific heuristics are and are not effective, i.e., to develop maps of the regions of rationality. The second, that awaits further research, is to elucidate the conditions under which people do or do not acquire such knowledge. In other words, how do people build maps of their decision making terrain? To be effective, people do not need much computational ability to make choices in the mazes that define their environments. However, they do need task-specific knowledge or maps.

References

Baucells, M., Carrasco, J. A., & Hogarth, R. M. (2006). *Cumulative dominance and heuristic performance in binary multiattribute choice.* Barcelona: Working paper IESE, UPC, and UPF.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, 25, 187-217.

Conlisk, J. (1996). Why bounded rationality? *Journal of Economic Literature*, 34 (2), 669-700.

Dawes, R. M., (1979). The robust beauty of improper linear models. *American Psychologist*, 34, 571-582.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95-106.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668-1674.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 13, 171-192.

Fasolo, B., McClelland, G. H., & Todd, P. M. (in press). Escaping the tyranny of choice: When fewer attributes make choice easier. *Marketing Theory*.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650-669.

Gigerenzer, G., & Selten, R. (2001). (Eds.) *Bounded rationality: The adaptive toolbox*. Cambridge, MA: The MIT Press.

Gigerenzer, G., Todd, P. M., and the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109 (1), 75-90.

Goldstein, W. M., & Hogarth, R. M. (1997). Judgment and decision research: Some historical context. In W. M. Goldstein & R. M. Hogarth (eds.), *Research on judgment and decision making: Currents, connections and controversies*. New York: Cambridge University Press.

Hertwig, R., & Todd, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. J. Hardman & L. Macchi (eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making*. Chichester, England: Wiley.

Hogarth, R. M., & Karelaia, N. (2004). *"Take-the-Best" and other simple strategies: Why and when they work "well" in binary choice*. Barcelona: Universitat Pompeu Fabra, DEE working paper.

Hogarth, R. M., & Karelaia, N. (2005a). Simple models for multi-attribute choice with many alternatives: When it does and does not pay to face tradeoffs with binary attributes. *Management Science*, 51 (12), 1860-1872.

Hogarth, R. M., & Karelaia, N. (2005b). Ignoring information in binary choice with continuous variables: When is less "more"? *Journal of Mathematical Psychology,* 49, 115-124.

Hogarth, R. M., & Karelaia, N. (2006). Online supplement for *Regions of rationality: Maps for bounded agents*. See www.xxxxxxxxxxxxxxxx.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5), 1449-1475.

Kahneman, D., Slovic, P., & Tversky, A. (1982). (Eds.) *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Karelaia, N. (in press). Thirst for confirmation in multi-attribute choice: Does search for consistency impair decision performance? *Organisational Behavior and Human Decision Processes*.

Karelaia, N., & Hogarth, R. M. (in preparation). *On predicting performance of DEBA models in the presence of error*. Barcelona: Universitat Pompeu Fabra

Katsikopoulos, K., & Martignon, L. (in press). Naïve heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology*.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, 107, 296-310.

Koehler, D., & Harvey, N. (eds.) (2004). *Blackwell Handbook of Judgment and Decision Making*. Oxford, UK: Blackwell.

Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the 'take the best' and 'rational' models. *Psychonomic Bulletin & Review,* 11, 343-352.

Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G Gigerenzer, P. M. Todd and the ABC Research Group. *Simple heuristics that make us smart* (pp. 119-140). New York: Oxford University Press.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52, 29-71.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.

Shanteau, J., & Thomas, R. P. (2000). Fast and frugal heuristics: What about unfriendly environments? *Behavioral and Brain Sciences*, 23, 762-763.

Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review*, 63, 129-138.

Simon, H. A. (1978). Rationality as process and product of thought. *American Economic Review*, 68 (2), 1-16

Simon, H. A. (1991). *Models of my life*. New York: Basic Books.

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science,* 25, 219-225.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281-299.

Table 1 -- Models tested

	Model	Description	Prior information*	Information to consult	Calculations	Number of comparisons min	Number of comparisons max
(A). Single Variable (SV) models							
1	Lexicographic -- SVc	Choice depends solely on cue with the greatest validity (see, e.g., Payne et al., 1993).	Rank-order of importance of cues	The most important cue	None	$(m-1)$	$\frac{m(m-1)}{2}$
2	Lexicographic -- SVb	Model 1 above but based on binary variables.					
3	DEBA	Deterministic version of Tversky's (1972) elimination-by-aspects model (EBA). For binary choice, this is the same as the take-the-best model of Gigerenzer and Goldstein (1996).		Variable			$k\frac{m(m-1)}{2}$
(B). Equal weight (EW) models							
4	EWc	All cues are accorded equal weight (see, e.g., Einhorn & Hogarth, 1975).	None	All	m sums	$(m-1)$	$\frac{m(m-1)}{2}$
5	EWb	All cues are accorded equal weight (see, e.g., Dawes, 1979),					
(C). Hybrid models							
6	EW/DEBA	Choose according to equal weights. If this results in a tie, use DEBA to resolve the choice (Hogarth & Karelaia, 2005a).	Rank-order of importance of cues	All	m sums	$(m-1)$	$\frac{m(m-1)}{2}$
7	EW/SVb	Choose according to equal weights. If this results in a tie, resolve conflict by the single most important variable					
(D). Domran (DR) models							
8	DRc	If an alternative dominates, choose it. Otherwise, choose at random between non-dominated alternatives.	None	All	None	$k(m-1)$	$k\frac{m(m-1)}{2}$
9	DRb	Same as DRc except based on binary variables.					
(E). Multiple regression (MR)							
10	MRc	Well-known statistical model.	Importance of cues	All	km products, m sums	$(m-1)$	$\frac{m(m-1)}{2}$
11	MRb	Same as MRc except based on binary variables.					

* For all models, the decision maker is assumed to know the sign of the zero order correlation between cues and the criterion.
NOTE: m = number of alternatives, k = number of attributes, cues

Table 2 -- Simulation data: parameters and overall predictive accuracy of models (% correct), choosing best of three

	Case A					Case B					Case C					Case D					
subcases	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
Experimental design																					
1 $\max(\rho_{yx_i}) - \min(\rho_{yx_i})$	0.3	0.4	0.5	0.6	0.7	0.3	0.4	0.5	0.6	0.7	0.1	0.3	0.3	0.3	0.4	0.0	0.1	0.3	0.4	0.5	
2 ρ_{yx_1}	0.4	0.5	0.6	0.7	0.8	0.4	0.5	0.6	0.7	0.8	0.3	0.4	0.5	0.5	0.6	0.3	0.4	0.6	0.7	0.8	
3 $\overline{\rho}_{yx_i}$	0.3	0.3	0.3	0.4	0.4	0.3	0.3	0.3	0.4	0.4	0.3	0.3	0.3	0.4	0.4	0.3	0.3	0.4	0.4	0.4	
4 $\overline{\rho}_{x_i x_j}$	0.0	0.0	0.0	0.0	0.0	0.5	0.5	0.5	0.5	0.5	0.1	0.1	0.1	0.1	0.1	0.6	0.6	0.6	0.6	0.6	
5 $\rho_{y\overline{x}}$	0.5	0.5	0.6	0.6	0.7	0.3	0.4	0.4	0.4	0.5	0.5	0.6	0.7	0.7	0.8	0.4	0.4	0.5	0.5	0.5	
6 R^2 (MR) -- fit	0.4	0.4	0.5	0.6	0.8	0.3	0.4	0.5	0.6	0.8	0.4	0.5	0.6	0.7	0.9	0.3	0.4	0.5	0.6	0.8	
7 (n-1)/(n-k)	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	
Results																					
Single Variable models																					Mean
1 Lexicographic -- SVc	50	54	**58**	**64**	**71**	**51**	**53**	**58**	**64**	**70**	47	52	53	55	62	46	49	**58**	**64**	**70**	57
2 Lexicographic -- SVb	44	47	50	53	57	44	47	50	52	56	43	44	47	48	52	41	44	51	54	57	49
3 DEBA	50	54	56	59	63	46	48	50	53	55	48	53	57	61	66	45	49	53	55	57	54
Equal Weight models																					
4 EWc	**52**	**55**	57	61	64	45	48	51	51	52	**52**	**57**	**62**	**66**	**74**	**47**	**50**	51	52	53	55
5 EWb	47	49	51	54	55	44	45	47	47	50	48	51	56	58	65	46	47	52	51	54	51
Hybrid models																					
6 EW/DEBA	50	53	54	58	62	46	47	49	52	53	50	54	58	61	68	**47**	49	50	51	54	53
7 EW/SVb	49	52	53	57	61	45	47	48	51	52	50	54	57	60	66	**47**	49	50	51	54	53
Domran models																					
8 DRc	41	42	43	44	45	42	44	46	47	48	37	37	38	39	39	43	44	45	46	48	43
9 DRb	47	47	49	51	53	43	46	47	48	50	43	46	48	51	53	45	47	48	49	52	48
Multiple regression																					
10 MRc	50	**55**	**59**	**66**	**74**	46	51	**58**	**66**	**75**	47	52	61	**67**	**79**	41	45	54	61	**71**	59
11 MRb	42	45	48	52	55	42	46	48	52	58	40	43	44	47	51	39	40	46	50	55	47
Mean	47	50	53	56	60	45	48	50	53	56	46	49	53	56	61	44	47	51	53	57	

Note: **Bold** figures denote largest realization in each column or second largest if MRc is largest.

Table 3-- Regressions of model performance on environmental characteristics

Models:	SVc	SVb	DEBA	EWc	EWb	EW/ DEBA	EW/ SVb	DRc	DRb	MRc	MRb
Regression coefficients* for:											
Constant	42	47	44	42	44	42	43	50	48	39	49
Dummy1	-12	-14	-13	-13	-14	-14	-13	-15	-14	-11	-13
Dummy2	-7	-9	-8	-7	-8	-8	-8	-8	-8	-6	-9
Number of cues					1	-1*		-5		-5	-2*
ρ_{yx_1}	50	29	11					28			
$\rho_{y\bar{x}}$			32	50	37	47	33	-16			-16
$\max(\rho_{yx_i}) - \min(\rho_{yx_i})$								-21			10*
$\overline{\rho}_{yx_i}$							19	34	46		
$\overline{\rho}_{x_i x_j}$					4	5			-4		
R^2	....	....	....	....	....	....	....	....	....	61	29
Regression statistics:											
R^2	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.98	0.99
Estimated standard error	1.30	0.71	1.13	1.43	1.04	1.13	1.23	0.64	0.73	1.93	0.92

* Significance level: $p < .05$. All other regression coefficients are statistically significant with $p < .001$.

Table 4-- Empirical datasets: parameters and overall predictive accuracy of models (% correct)

Experimental design

		Golf rankings	Golf earnings	PhD programs	Digital cameras
1	$\max(\rho_{yx_i}) - \min(\rho_{yx_i})$	0.2	0.3	0.1	0.4
2	ρ_{yx_1}	0.8	0.9	0.8	0.8
3	$\bar{\rho}_{yx_i}$	0.6	0.7	0.8	0.5
4	$\bar{\rho}_{x_i x_j}$	0.5	0.5	0.6	0.2
5	$\rho_{y\bar{x}}$	0.8	0.8	0.9	0.8
6	R^2 (MR) -- fit	0.7	0.8	0.8	0.7
7	(n-1)/(n-k)	1.07	1.07	1.04	1.09

Results

		Golf rankings			Golf earnings			PhD programs			Digital cameras			
		Best of			Best of			Best of			Best of			
		Two	Three	Four	Two	Three	Four	Two	Three	Four	Two	Three	Four	Mean
Single Variable models														
1	Lexicographic -- SVc	**79**	**72**	**67**	**81**	**76**	**73**	78	75	74	72	66	59	73
2	Lexicographic -- SVb	69	56	46	73	61	53	71	63	57	69	64	58	62
3	DEBA	73	60	49	77	71	65	80	76	74	**79**	**73**	**67**	70
Equal Weight models														
4	EWc	77	68	62	77	72	68	**86**	**83**	**81**	76	72	65	74
5	EWb	71	58	48	75	67	63	79	75	73	77	69	63	68
Hybrid models														
6	EW/DEBA	73	59	49	77	71	65	79	77	74	**79**	71	66	70
7	EW/SVb	72	58	49	78	71	65	79	76	74	**79**	71	66	70
Domran models														
8	DRc	69	56	47	69	61	57	77	69	66	73	62	55	63
9	DRb	69	57	48	74	65	59	78	75	72	75	64	56	66
Multiple regression														
10	MRc	**80**	71	65	**82**	**78**	**75**	**86**	**83**	**81**	78	72	66	76
11	MRb	67	54	47	71	62	54	77	73	69	69	66	59	64
	Mean	73	61	53	76	69	63	79	75	72	75	68	62	

Note: **Bold** figures denote largest realization in each column or second largest if MRc is largest.

Figure 1

Percentage correct predictions by different models for conditions specified in Table 2 (cases A, B, C, and D): choosing one of three.

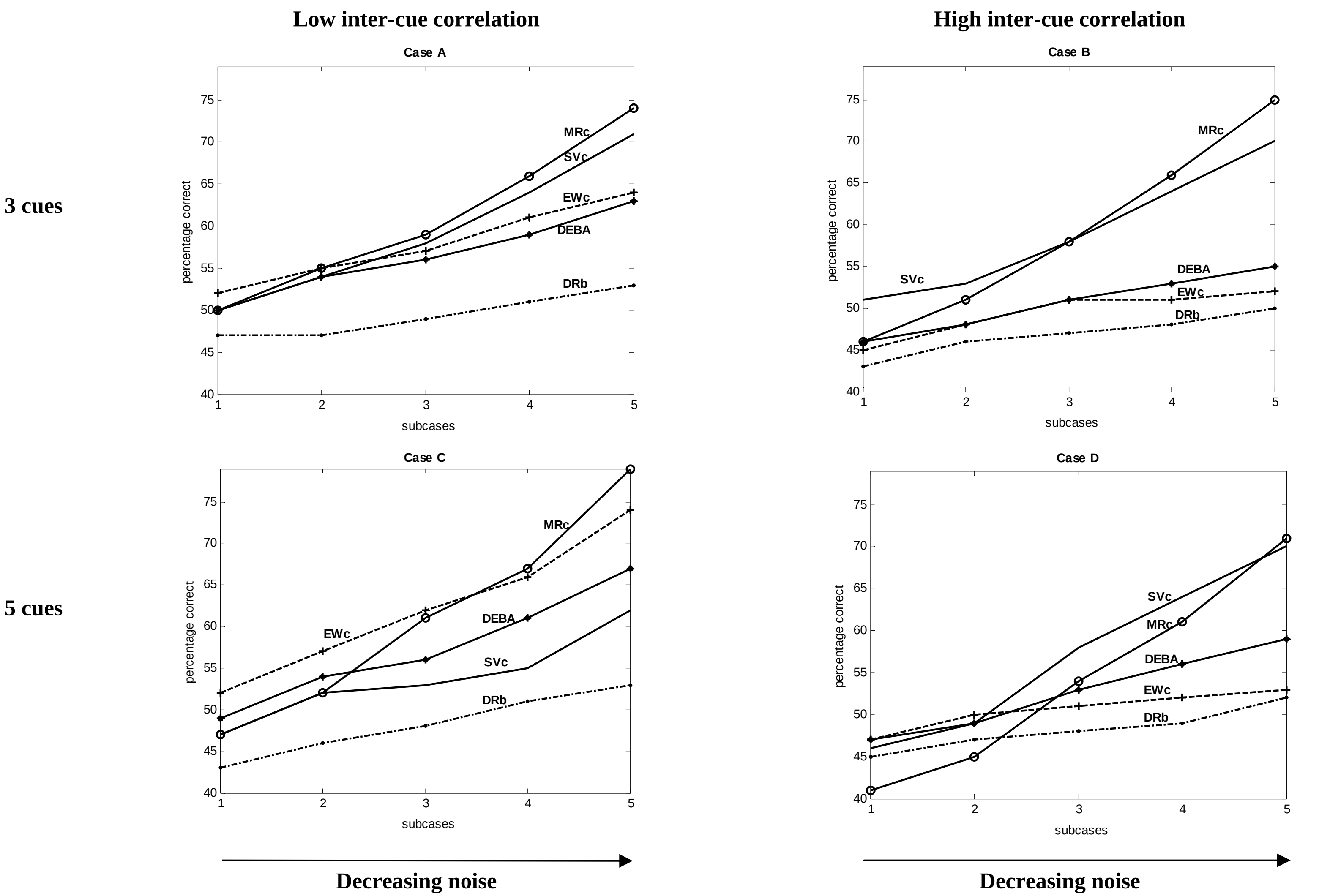
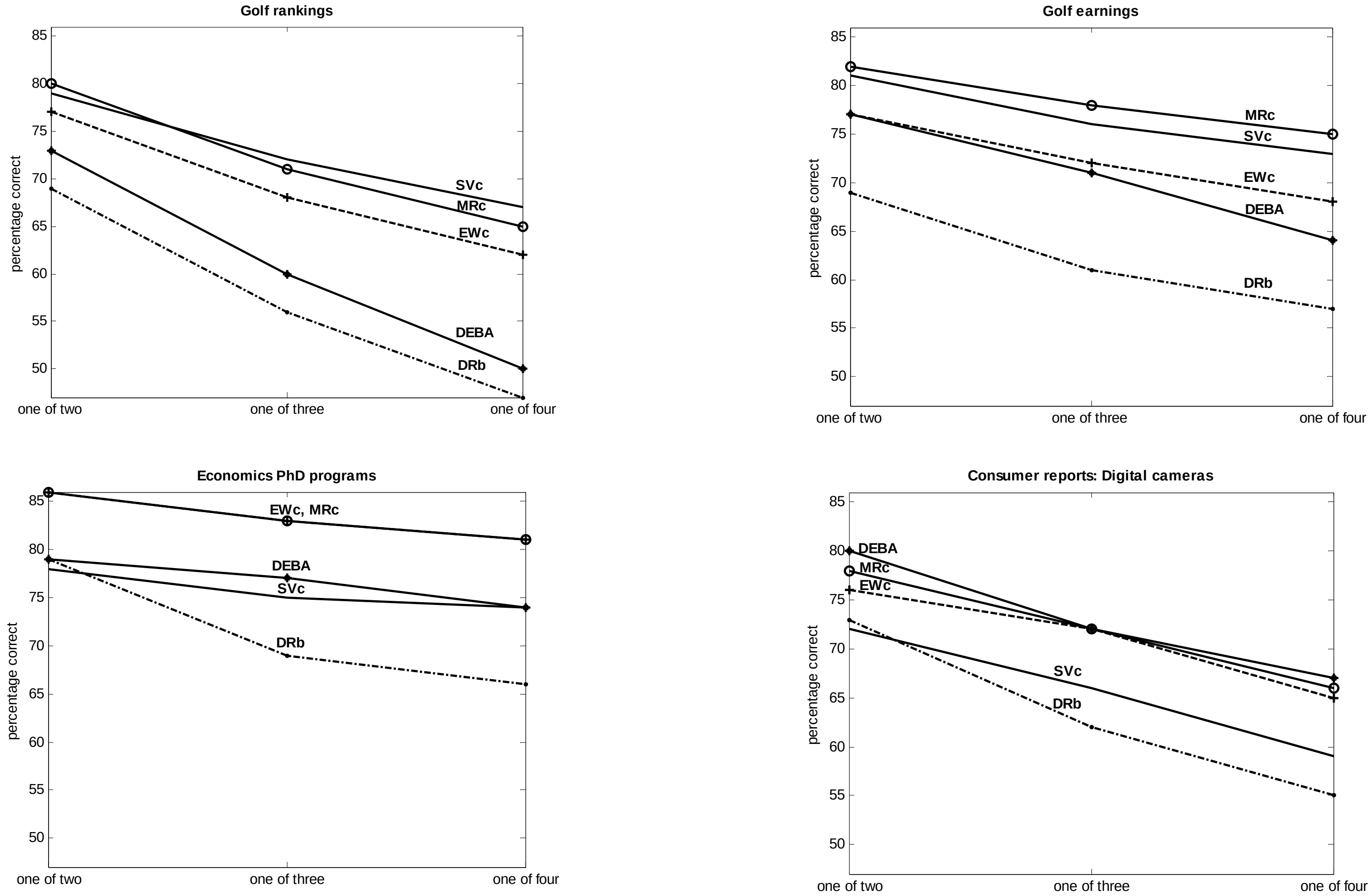


Figure 2
Percentage correct predictions by different models for real data sets specified in Table 4.



Appendix A – Key formulas for different models using continuous variables (for choosing best of m)

Model (for alternative i)	Error variance	Upper limits of integration d_t^*, for $t=\overline{1,m-1}$	d_t	V_d
Single variable (SVc) $Y_i=\rho_{yx}X_i+\varepsilon_i$	$2(1-\rho_{yx}^2)$	$\frac{\rho_{yx}d_t}{\sqrt{2\left(1-\rho_{yx}{}^2\right)}}$	x_i-x_j	$\begin{pmatrix} 2 & \dots & 1 \\ \dots & \dots & \dots \\ 1 & \dots & 2 \end{pmatrix}$
Equal weights (EWc) $Y_i=\frac{\rho_{y\bar{x}}}{\sigma_{\bar{x}}}\overline{X}_i+v_i$	$2(1-\rho_{y\bar{x}}^2)$	$\frac{\rho_{y\bar{x}}d_t}{\sigma_{\bar{x}}\sqrt{2\left(1-\rho_{y\bar{x}}{}^2\right)}}$	$\bar{x}_i-\bar{x}_j$	$\begin{pmatrix} \sqrt{2}\sigma_{\bar{x}} & \dots & \sigma_{\bar{x}}^2 \\ \dots & \dots & \dots \\ \sigma_{\bar{x}}^2 & \dots & \sqrt{2}\sigma_{\bar{x}} \end{pmatrix}$
Multiple regression (MRc) $Y_i=\hat{Y}_i+u_i$	$2\,(1-R_{adj}^2)$	$\frac{d_t}{\sqrt{2(1-R_{adj}^2)}}$	$\hat{y}_i-\hat{y}_j$	$\begin{pmatrix} \sqrt{2}R_{adj} & \dots & R_{adj}^2 \\ \dots & \dots & \dots \\ R_{adj}^2 & \dots & \sqrt{2}R_{adj} \end{pmatrix}$

Notes:

1. Predictive accuracy of a single choice between m alternatives is given by

$$\int_{-\infty}^{d_1^*}\dots\int_{-\infty}^{d_{m-1}^*}\varphi(z|\mu_z,V_z)dz_1\dots.dz_{m-1},$$

where the pdf $\varphi(z|\mu_z,V_z)$ is defined for $z'=(z_1,z_2,\dots,z_{m-1})$ and $V_z=\begin{pmatrix} 1 & \dots & ½ \\ \dots & \dots & \dots \\ ½ & \dots & 1 \end{pmatrix}$,

d_t^* and d_t $(t=\overline{1,m-1})$ being specific for different choice strategies.

2. Overall predictive accuracy of choice (between m alternatives) in a given population is given by

$$m\int_0^{\infty}\dots.\int_0^{\infty}\varphi(d|\mu_d,V_d)\left[\int_{-\infty}^{d_1^*}\dots.\int_{-\infty}^{d_{m-1}^*}\varphi(z|\mu_z,V_z)dz_1\dots dz_{m-1}\right]dd_1\dots dd_{m-1},$$

where the pdf $\varphi(d|\mu_d,V_d)$ is defined for $d'=(d_1,d_2,\dots,d_{m-1})$, V_d being specific for different choice strategies.

3. $\rho_{y\bar{x}}=\bar{\rho}_{yx}\sqrt{\frac{k}{1+(k-1)\bar{\rho}_{x_ix_j}}}$, where k = number of x variables, $\bar{\rho}_{yx}$ = average correlation between y and the x's, and $\bar{\rho}_{x_ix_j}$ = average inter-correlations amongst the x's.

4. $\sigma_{\bar{x}}=\sqrt{\frac{1}{k}\left(1+(k-1)\bar{\rho}_{x_ix_j}\right)}$

5. $R_{adj}^2=1-\left(1-R^2\right)\frac{(n-1)}{(n-k)}$, where n = number of observations.

Appendix B – Key formulas for different models using binary variables (for choosing best of *m*)

Model (for alternative *i*)	Error variance	Upper limits of integration d_t^*, for $t=\overline{1,m-1}$	d_t	V_d
Single variable (SVb) $Y_i = a_{SVb} + 2\rho_{yw}W_i + \varsigma_i$	$2(1-\rho_{yw}^2)$	$\dfrac{2\rho_{yw}d_t}{\sqrt{2\left(1-\rho_{yw}^2\right)}}$	$w_i - w_j$	$\begin{pmatrix} \frac{1}{2} & \dots & \frac{1}{4} \\ \dots & \dots & \dots \\ \frac{1}{4} & \dots & \frac{1}{2} \end{pmatrix}$
Equal weights (EWb) $Y_i = a_{EWb} + \dfrac{\rho_{y\overline{w}}}{\sigma_{\overline{w}}}\overline{W}_i + \xi_i$	$2(1-\rho_{y\overline{w}}^2)$	$\dfrac{\rho_{y\overline{w}}d_t}{\sigma_{\overline{w}}\sqrt{2\left(1-\rho_{y\overline{w}}^2\right)}}$	$\overline{w}_i - \overline{w}_j$	$\begin{pmatrix} \sqrt{2}\sigma_{\overline{w}} & \dots & \sigma_{\overline{w}}^2 \\ \dots & \dots & \dots \\ \sigma_{\overline{w}}^2 & \dots & \sqrt{2}\sigma_{\overline{w}} \end{pmatrix}$
Multiple regression (MRb) $Y_i = \hat{Y}_i + \zeta_i$	$2(1-R_{adj}^2)$	$\dfrac{d_t}{\sqrt{2(1-R_{adj}^2)}}$	$\hat{y}_i - \hat{y}_j$	$\begin{pmatrix} \sqrt{2}R_{adj} & \dots & R_{adj}^2 \\ \dots & \dots & \dots \\ R_{adj}^2 & \dots & \sqrt{2}R_{adj} \end{pmatrix}$

Notes:

1. $\hat{Y}$ and R_{adj}^2 are based on binary variables *W*'s.
2. Since the error terms have means of zero, the intercepts a_{SVb} and a_{EWb} are equal to $-\rho_{yw}$ and $-\dfrac{\rho_{y\overline{w}}}{2\sigma_{\overline{w}}}$, respectively.
3. General formulas for predictive accuracy of a single choice and overall predictive accuracy are as specified in Notes 1 and 2 to Appendix A.
4. $\rho_{y\overline{w}} = \overline{\rho}_{yw}\sqrt{\dfrac{k}{1+(k-1)\overline{\rho}_{w_iw_j}}}$, where *k* = number of *x* variables, $\overline{\rho}_{yw}$ = average correlation between *y* and the *w*'s, and $\overline{\rho}_{w_iw_j}$ = average inter-correlations amongst the *w*'s.
5. $\sigma_{\overline{w}} = \sqrt{\dfrac{1}{k}\left(1+(k-1)\overline{\rho}_{w_iw_j}\right)}$

Appendix C – DEBA model: Two examples of probability calculations

Example 1. Assume that there are three alternatives A, B, and C and that A has been chosen by a process whereby C was eliminated at the first stage and B at the third stage. Starting backwards, consider the decisions the model makes at each stage. That is, the probability that DEBA correctly selected A over B at the third stage, controlling for the elimination of C at the first stage, is $P\{(Y_a > Y_b | W_{a3} = w_{a3} > W_{b3} = w_{b3}) \cap (Y_b > Y_c | W_{b1} = w_{b1} > W_{c1} = w_{c1})\}$. This probability can be calculated by making use of the appropriate partial correlations – in this case, $\rho_{yw_3.w_1w_2}$ and ρ_{yw_1} – and adapting the single variable equations (e.g., the general equation 6 [24]). At the second stage, the model makes no decision. At the first stage, it eliminates C so we need to calculate additionally the probability that A could have been correctly selected only with information available at this stage: $P\{(Y_a > Y_c | W_{a1} = w_{a1} > W_{c1} = w_{c1}) \cap (Y_c > Y_b | W_{b1} = w_{b1} > W_{c1} = w_{c1})\}$. This can be also found through an adapted expression (6), using ρ_{yw_1}. Importantly, the events represented by the probability expressions for the first and third stages are disjunctive. Therefore, the probability that DEBA makes the correct decision in this case is equal to the sum of the two expressions.

Example 2. Assume that with 3 other alternatives A, B, and C, DEBA eliminates C at the first stage and at the third stage picks either A or B at random (this will happen if A and B are identical). Thus, the 0.5 probability that DEBA makes the correct decision at the third stage should be "discounted" by the probability that C, eliminated at the first stage, is not better than A and B. That is $0.5(1 - P\{(Y_c > Y_a | W_{c1} = w_{c1} > W_{a1} = w_{a1}) \cap (Y_c > Y_b | W_{c1} = w_{c1} > W_{b1} = w_{b1})\})$.

[24] The terms that need to be adapted in the expression (6) are the upper limits of integration and σ_{z_1, z_2}.