

Heuristic and Linear Models of Judgment: Matching Rules and Environments

Robin M. Hogarth
ICREA and Universitat Pompeu Fabra

Natalia Karelaia
University of Lausanne

Much research has highlighted incoherent implications of judgmental heuristics, yet other findings have demonstrated high correspondence between predictions and outcomes. At the same time, judgment has been well modeled in the form of *as if* linear models. Accepting the probabilistic nature of the environment, the authors use statistical tools to model how the performance of heuristic rules varies as a function of environmental characteristics. They further characterize the human use of linear models by exploring effects of different levels of cognitive ability. They illustrate with both theoretical analyses and simulations. Results are linked to the empirical literature by a meta-analysis of lens model studies. Using the same tasks, the authors estimate the performance of both heuristics and humans where the latter are assumed to use linear models. Their results emphasize that judgmental accuracy depends on matching characteristics of rules and environments and highlight the trade-off between using linear models and heuristics. Whereas the former can be cognitively demanding, the latter are simple to implement. However, heuristics require knowledge to indicate when they should be used.

Keywords: decision making, heuristics, linear models, lens model, judgmental biases

Two classes of models have dominated research on judgment and decision making over past decades. In one, explicit recognition is given to the limits of information processing, and people are modeled as using simplifying heuristics (Gigerenzer, Todd, & the ABC Research Group, 1999; Kahneman, Slovic, & Tversky, 1982). In the other, it is assumed that people can integrate all the information at hand and that this is combined and weighted as if using an algebraic—typically linear—model (Anderson, 1981; Brehmer, 1994; Hammond, 1996).

Robin M. Hogarth, ICREA and Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain; Natalia Karelaia, Department of Management, University of Lausanne, Lausanne, Switzerland.

The authors' names are listed alphabetically.

This research has been funded within the EUROCORES European Collaborative Research Project (ECRP) Scheme jointly provided by ECRP funding agencies and the European Science Foundation. A list of ECRP funding agencies can be consulted at http://www.esf.org/ecrp_countries. It was specifically supported by Spanish Ministerio de Educación y Ciencia Grant SEC2005-25690 (Hogarth) and Swiss National Science Foundation Grant 105511-111621 (Karelaia). We have greatly benefited from the comments of Michael Doherty, Joshua Klayman, and Chris White, as well as from presentations at the annual meeting of the Brunswik Society, Toronto, Ontario, Canada, November 2005; FUR XII, Rome, Italy, June 2006; the University of Basel, Basel, Switzerland; INSEAD, Fontainebleau, France; and the Max Planck Institute, Berlin, Germany. We are particularly indebted to Thomas Stewart, Michael Doherty, and the library at Universitat Pompeu Fabra for helping us locate many lens model studies, as well as to Marcus O'Connor for providing data.

Correspondence concerning this article should be addressed to Robin M. Hogarth, Universitat Pompeu Fabra, Department of Economics and Business, Ramon Trias Fargas 25-27, 08005, Barcelona, Spain, or to Natalia Karelaia, Department of Management, University of Lausanne, Internef, 1015, Lausanne-Dorigny, Switzerland. E-mail: robin.hogarth@upf.edu or natalia.karelaia@unil.ch

The topic of heuristics has generated many interesting findings, as well as controversy (see, e.g., Gigerenzer, 1996; Kahneman & Tversky, 1996). However, whereas few scholars doubt that people make extensive use of heuristics (as variously defined), many questions are unresolved. One important issue—and key to the controversy—has been the failure to explicate the relative efficacy of heuristics and especially to define a priori the environmental conditions when these are differentially accurate.

At one level, this failure is surprising in that Herbert Simon—whose work is held in high esteem by researchers with opposing views about heuristics—specifically emphasized environmental factors. Indeed, some 50 years ago, Simon stated,

if an organism is confronted with the problem of behaving approximately rationally, or adaptively, in a particular environment, the kinds of simplifications that are suitable may depend not only on the characteristics—sensory, neural, and other—of the organism, but equally on the nature of the environment. (Simon, 1956, p. 130)

At the same time that Simon was publishing his seminal work on bounded rationality, the use of linear models to represent psychological processes received considerable impetus from Hammond's (1955) formulation of clinical judgment and was subsequently bolstered by Hoffman's (1960) argument for "paramorphic" representation (see also Einhorn, Kleinmuntz, & Kleinmuntz, 1979). Contrary to work on heuristics, this research has shown concern for environmental factors.

Specifically—as illustrated in Figure 1—Hammond and his colleagues (Hammond, Hursch, & Todd, 1964; Hursch, Hammond, & Hursch, 1964; Tucker, 1964) depicted Brunswik's (1952) lens model within a linear framework that defines both judgments and the criterion being judged as functions of cues in the environment. Thus, the accuracy of judgment (or psychological achievement) depends on both the inherent predictability of the environment and the extent to which the weights

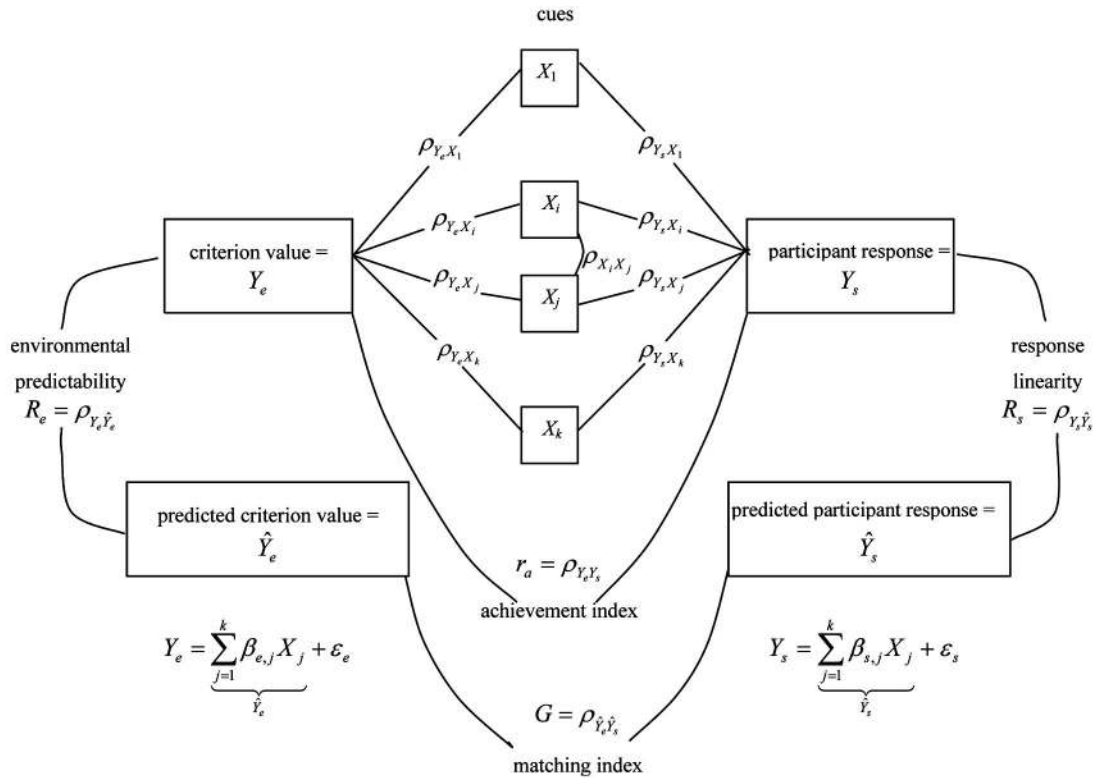


Figure 1. Diagram of lens model.

humans attach to different cues match those of the environment. In other words, accuracy depends on the characteristics of the cognitive strategies that people use and those of the environment. Moreover, this framework has been profitably used by many researchers (see, e.g., Brehmer & Joyce, 1988; Cooksey, 1996; Hastie & Kameda, 2005). Other techniques, such as conjoint analysis (cf. Louvière, 1988), also assume that people process information as though using linear models and, in so doing, seek to quantify the relative weights given to different variables affecting judgments and decisions (see also Anderson, 1981).

In many ways, the linear model has been the workhorse of judgment and decision-making research from both descriptive and prescriptive viewpoints. As to the latter, consider the influence of linear models on decision analysis (see, e.g., Keeney & Raiffa, 1976), prediction tasks (Camerer, 1981; Dawes, 1979; Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Goldberg, 1970; Wainer, 1976), and the statistical-clinical debate (Dawes, Faust, & Meehl, 1989; Kleinmuntz, 1990; Meehl, 1954).

Despite the ubiquity of the linear model in representing human judgment, its psychological validity has been questioned for many decision-making tasks. First, when the amount of information increases (e.g., more than three cues in a multiple-cue prediction task), people have difficulty in executing linear rules and resort to simplifying heuristics. Second, the linear model implies trade-offs between cues or attributes, and because people find these difficult to execute—both cognitively and emotionally (Hogarth, 1987; Luce, Payne, & Bettman, 1999)—they often resort to trade-off-

avoiding heuristics (Montgomery, 1983; Payne, Bettman, & Johnson, 1993).

This discussion of heuristics and linear models raises many important psychological issues. Under what conditions do people use heuristics—and which heuristics—and how accurate are these relative to the linear model? Moreover, if heuristics neglect information and/or avoid trade-offs, how do these features contribute to their success or failure, and when?

A further issue relates to how heuristic performance is evaluated. One approach is to identify instances in which heuristics violate coherence with the implications of statistical theory (see, e.g., Tversky & Kahneman, 1983). The other considers the extent to which predictions match empirical realizations (Gigerenzer et al., 1999). These two approaches, labeled coherence and correspondence, respectively (Hammond, 1996), may sometimes conflict in the impressions they imply of people's judgmental abilities. In this article, we follow the second because our goal is to understand how the performance of heuristic rules and linear models is affected by the characteristics of the environments in which they are used. In other words, we speak directly to the need specified by Simon (1956) to develop a theory of how environmental characteristics affect judgment (see also Brunswik, 1952).

This article is organized as follows. We first outline the framework within which our analysis is conducted and specify the particular models used in our work. We then briefly review literature that has considered the accuracy of heuristic decision models. For the most part, this has involved empirical demonstrations and simulations, and thus, conclusions cannot be easily generalized. In

contrast, our approach, developed in the subsequent section, explicitly recognizes the probabilistic nature of the environment and exploits appropriate statistical theory. This allows us to make theoretical predictions of model accuracy in terms of both percentage of correct predictions and expected losses. We emphasize here that these predictions are theoretical implications, as opposed to forecasts made by fitting models to data and extrapolating to new samples. Briefly, the rationale for this approach—discussed further below—is to capture the power of theory to make claims that can be generalized. To facilitate the exposition, we do not present the underlying rationales for all models in the main text but make use of appendices. We demonstrate the power of our equations with theoretical predictions of differential model performance over a wide range of environments, as well as using simulation. This is followed by our examination of empirical data using a meta-analysis of the lens model literature. Finally, we consider psychological, normative, and methodological implications of our work, as well as suggestions for future research.

Framework and Models

We conduct our analyses within the context of predicting (choosing) the better of two alternatives on the basis of several cues (attributes). Moreover, we assume that the criterion is probabilistically related to the cues and that the optimal equation for predicting the criterion is a linear function of the cues.¹ Thus, if the decision maker weights the cues appropriately (using a linear model), he or she will achieve the maximum predictive performance. However, this could be an exacting standard to achieve. Thus, what are the consequences of abandoning the linear rule and using simpler heuristics? Moreover, when do different heuristics perform relatively well or badly?

Specifically, we consider five models and, to simplify the analysis, consider only three cues. Two of these models are linear, and three are heuristics. Whereas we could have chosen many variations of these models, they are sufficient to illustrate our approach.

First, we consider what happens when the decision maker can be modeled as if he or she were using a linear combination (LC) of the cues but is inconsistent (cf. Hoffman, 1960). Note carefully that we are not saying that the decision maker actually uses a linear formula but that this can be modeled *as if*. We justify this on the grounds that linear models can often provide higher level representations of underlying processes (Einhorn et al., 1979). Moreover, when the information to be integrated is limited, the linear model can also provide a good process description (Payne et al., 1993).

Second, the decision maker uses a simplified version of the linear model that gives equal weight (EW) to all variables (Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975).²

Third, the decision maker uses the take-the-best (TTB) heuristic proposed by Gigerenzer and Goldstein (1996). This model first assumes that the decision maker can order cues or attributes by their ability to predict the criterion. Choice is then made by the most predictive cue that can discriminate between options. If no cues discriminate, choice is made at random. This model is “fast and frugal” in that it typically decides on the basis of one or two cues (Gigerenzer et al., 1999).³

There is experimental evidence that people use TTB-like strategies, although not exclusively (Bröder, 2000, 2003; Bröder & Schiffer, 2003; Newell & Shanks, 2003; Newell, Weston, &

Shanks, 2003; Rieskamp & Hoffrage, 1999). Descriptively, the two most important criticisms are, first, that the stopping rule is often violated in that people seek more information than the model specifies and, second, that people may not be able to rank-order the cues by predictive ability (Juslin & Persson, 2002).

The fourth model, CONF (Karelaia, 2006), was developed to overcome the descriptive shortcomings of TTB. Its spirit is to consult the cues in the order of their validity (like TTB) but not to stop the process once a discriminating cue has been identified. Instead, the process only stops once the discrimination has been confirmed by another cue.⁴ With three cues, then, CONF requires only that two cues favor the chosen alternative. Moreover, CONF has the advantage that choice is insensitive to the order in which cues are consulted. The decision maker does not need to know the relative validities of the cues.⁵

Finally, our fifth model is based solely on the single variable (SV) that the decision maker believes to be most predictive. Thus, this differs from TTB in that, across a series of judgments, only one cue is consulted. Parenthetically, this could also be used to model any heuristic based on one variable, such as judgments by availability (Tversky & Kahneman, 1973), recognition (Goldstein & Gigerenzer, 2002), or affect (Slovic, Finucane, Peters, & MacGregor, 2002). In these cases, however, the variable would not be a cue that could be observed by a third party but would represent an intuitive feeling or judgment experienced by the decision maker (e.g., ease of recall, sensation of recognition, or a feeling of liking).

It is important to note that all these rules represent feasible psychological processes. Table 1 specifies and compares what needs to be known for each of the models to achieve its maximum performance. This can be decomposed between knowledge about the specific cue values (on the left) and what is needed to weight the variables (on the right). Two models require knowing all cue values (LC and EW), and one only needs to know one (SV). The number of cue values required by TTB and CONF depends on the characteristics of each choice faced. As to weights, maximum performance by LC requires precise, absolute knowledge; TTB requires the ability to rank-order cues by validity; and for SV, one

¹ Whereas the linear assumption is a limitation, we note that many studies have shown that linear functions can approximate nonlinear functions well provided the relations between the cues and criterion are conditionally monotonic (see, e.g., Dawes & Corrigan, 1974).

² In all of the models investigated, we assume that if the decision maker uses a variable, he or she knows its zero-order correlation with the criterion.

³ In Gigerenzer and Goldstein's (1996) formulation, TTB operates only on cues that can take binary values (i.e., 0/1). We analyze a version of this model based on continuous cues where discrimination is determined by a threshold, that is, a cue discriminates between two alternatives only if the difference between the values of the cues exceeds a specified value $t (>0)$.

⁴ In our subsequent modeling of CONF, we assume that any difference between cue values is sufficient to indicate discrimination or confirmation. In principle, one could also assume a threshold in the same way that we model TTB.

⁵ Parenthetically, with $k > 3$ cues, CONF is also insensitive to cue ordering as long as the model requires at least $k/2$ confirming cues when k is even and at least $(k - 1)/2$ confirming cues when k is odd (Karelaia, 2006).

Table 1
Knowledge Required to Achieve Upper Limits of Model Performance

Model	Values of variables ^a			Weights ordering			
	Cue 1	Cue 2	Cue 3	Exact ^b	First ^c	All ^d	None
Linear combination (LC)	Yes	Yes	Yes	Yes			
Equal weighting (EW)	Yes	Yes	Yes				Yes
Take-the-best (TTB)	Yes	Yes/no	Yes/no			Yes	
Single variable (SV)	Yes				Yes		
CONF	Yes	Yes	Yes/no				Yes

^a Yes = value of cue required; Yes/no = value of cue may be required.

^b Exact values of cue weights required.

^c First = most important cue identified.

^d All = rank order of all cues known a priori.

needs to identify the cue with the greatest validity. Neither EW nor CONF requires knowledge about weights.

Whereas it is difficult to tell whether obtaining values of cue variables or knowing how cues vary in importance is more taxing cognitively, we have attempted an ordering of the models in Table 1 from most to least taxing. LC is the most taxing *ceteris paribus*. The important issue is to characterize its sensitivity to deviations from optimal specification of its parameters. CONF, at the other extreme, is not demanding, and the only uncertainty centers on how many variables need to be consulted for each decision.

In our analysis, we adopt a Brunswikian perspective by exploiting properties of the well-known lens model equation (Hammond et al., 1964; Hammond & Stewart, 2001; Hursch et al., 1964; Tucker, 1964). We combine this with more recent analytic methods developed to determine the performance of heuristic decision rules (Hogarth & Karelaia, 2005a, 2006a; Karelaia, 2006). Using these tools, we are able to describe how environmental characteristics interact with those of the different heuristics in determining the performance of the latter.

The novelty of our approach is that we are able to compare and contrast heuristic and linear model performance within the same analytical framework. Moreover, noting that different models require different levels of knowledge (see Table 1), we see our work as specifying the demand for knowledge in different regions of the environment. In other words, to make accurate decisions, how much and what knowledge is needed in different types of situations?

In brief, our analytical results show that the performance of heuristic rules is affected by how the environment weights cues, cue redundancy, the predictability of the environment, and loss functions. Heuristics predict accurately when their characteristics match the demands of the environment; for example, EW is best when the environment also weights the cues equally. However, in the absence of a close match between characteristics of heuristics and the environment, the presence of redundancy can moderate the relative predictive ability of different heuristics. Both cue redundancy and noise (i.e., lack of predictability) also reduce differences between model performances, but these can be augmented or diminished according to the loss function. We also show that sensible models often make identical predictions. However, because they disagree across 8%–30% of the cases we examine, it pays to understand the differences.

We exploit the mathematics of the lens model (Tucker, 1964) to ask how well decision makers need to execute LC rule strategies to perform as well as or better than heuristics in binary choice using the criterion of predictive accuracy (i.e., correspondence). We find that performance using LC rules generally falls short of that of appropriate heuristics unless decision makers have high *linear cognitive ability*, or *ca* (which we quantify). This analysis is supported by a meta-analysis of lens model studies in which we estimate *ca* across 270 tasks and also demonstrate that, within the same tasks, individuals vary in their ability to outperform heuristics using LC models. Finally, we illustrate how errors in the application of both linear models and heuristics affect performance and thus the nature of potential trade-offs involved in using different models.

Evidence on the Accuracy of Simple, Heuristic Models

Interest in the use of heuristics has fueled much research (and controversy) in judgment and decision making. Simon's work on bounded rationality (Simon, 1955, 1956) emphasized the need for humans to use heuristic methods (or to "satisfice") because of inherent cognitive limitations. Moreover, as noted above, Simon stressed the importance of understanding how the structure of the environment affects the performance of these heuristics.

This environmental concern, however, was largely lacking from the influential research on heuristics and biases spearheaded by Tversky and Kahneman (1974; see also Kahneman et al., 1982). As stated by these researchers, "these heuristics are highly economical and usually effective, but they lead to systematic and predictable errors" (Tversky & Kahneman, 1974, p. 1131). Unfortunately, no environmental theory was offered specifying the conditions under which heuristics are or are not accurate (see also Hogarth, 1981). Instead, the argument rested on demonstrating that some responses did not cohere with the dictates of statistical theory.

Nonetheless, the positive side of heuristic use has also been emphasized. One line of research has emphasized EW models, the accuracy of which was demonstrated through simulations and empirical examples (Dawes, 1979; Dawes & Corrigan, 1974). In further simulations, Payne et al. (1993) explored trade-offs between effort and accuracy. Using continuous variables and a weighted additive model as the criterion, they investigated the

performance of several simple choice strategies and specifically 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). Also using simulations, McKenzie (1994) showed how simple strategies of covariation judgment and Bayesian inference can achieve impressive performance.

The predictive accuracy of TTB was first demonstrated by Gigerenzer and Goldstein (1996) in an empirical illustration and subsequently replicated over 18 further data sets (Gigerenzer et al., 1999). Specifically, these studies showed that TTB predicts more accurately (on cross-validation) than EW and multiple regression when the criterion is the percentage of correct predictions (in binary choice). However, there was little concern as to whether these outcomes were the result of favorable environmental conditions. Voicing these concerns, Shanteau and Thomas (2000) constructed environments that they reasoned would be “friendly” or “unfriendly” to different models and demonstrated these effects through simulations. However, they did not address the issue of the relative frequencies of friendly and unfriendly environments in natural decision-making contexts.

Environmental effects were also demonstrated by Fasolo, McClelland, and Todd (2007) in a simulation of multiattribute choice using continuous variables (involving 21 options characterized by six attributes). 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 intercorrelations among the attributes and types of weighting functions. Results showed important effects for both. When true utility involved differential weighting, the most important attribute captured at least 90% of total utility. With positive intercorrelation among attributes, there was little difference between equal and differential weighting. With negative intercorrelation, however, equal weighting was sensitive to the number of attributes used (the more, the better).

Despite these empirical demonstrations involving simulated and real data, there has been relatively little theoretical work aimed at elucidating the environmental conditions under which heuristic models are and are not accurate. This is an important gap in scientific knowledge. That is, scientists know that various heuristics have been successful in some environments, but they do not know why and the extent to which results might generalize to other environments.

Some work has, however, considered specific cases. Einhorn and Hogarth (1975), for example, developed a theoretical rationale for the accuracy of EW relative to multiple regression. Klayman and Ha (1987) provided an illuminating account of why the so-called positive-test heuristic is highly effective when testing hypotheses in many types of environments. Martignon and Hoffrage (1999, 2002) and Katsikopoulos and Martignon (2006) explored the conditions under which TTB or EW should be preferred in binary choice. Hogarth and Karelaia (2005b, 2006b) and Baucells, Carrasco, and Hogarth (in press) have examined why TTB and other simple models perform well with binary attributes in error-free environments. Finally, in related work (Hogarth & Karelaia, 2005a, 2006a), we have provided an analytical framework for determining what we named *regions of rationality*, that is, the identification of environmental and model characteristics that

specify when heuristics do and do not predict accurately. The current article builds on these foundations.

The next section is technical. We first briefly explain the logic of the lens model and the lens model equation (Tucker, 1964). We then derive equations for the predictive ability of the heuristics we examine in terms of expected proportion of correct predictions in binary choice as well as squared-error loss functions. An important difference between studies of heuristic judgment and those using the lens model is that the empirical criterion for the latter—known as *achievement*—is measured by the correlation between judgments and outcomes as opposed to percentage of correct predictions in binary choice. To compare paradigms, we transform correlational achievement into equivalent percentage correct in binary choice.

Theoretical Development

Accepting the probabilistic nature of the environment (Brunswick, 1952), we use statistical theory to model both how people make judgments and the characteristics of the environments in which those judgments are made. To motivate the theoretical development, imagine a binary choice that involves selecting one of two job candidates, A and B, on the basis of several characteristics such as level of professional qualifications, years of experience, and so on. Further, imagine that a criterion can be observed at a later date and that a correct decision has been taken if the criterion is greater for the chosen candidate. Denote the criterion by the random variable Y_e such that if A happened to be the correct choice, one would observe $y_{ea} > y_{eb}$.⁶

Within the lens model framework—see Figure 1—we can model assessments of candidates by two equations: one, the model of the environment; the other, the model of the judge (the person assessing the job candidates). That is,

$$Y_e = \sum_{j=1}^k \beta_{e,j} X_j + \varepsilon_e, \quad (1)$$

and

$$Y_s = \sum_{j=1}^k \beta_{s,j} X_j + \varepsilon_s, \quad (2)$$

where Y_e represents the criterion (subsequent job performance of candidates) and Y_s is the judgment made by the decision maker, the X_j s are cues (here, characteristics of the candidates), and ε_e and ε_s are normally distributed error terms with means of zero and constant variances independent of the X s.

The logic of the lens model is that the judge's decisions will match the environmental criterion to the extent that the weights the judge gives to the cues match those used by the model of the environment, that is, the matches between $\beta_{s,j}$ and $\beta_{e,j}$ for all $j = 1, \dots, k$ —see Figure 1.

⁶ We use uppercase letters to denote random variables, for example, Y_e , and lowercase letters to designate specific values, for example, y_e . As exceptions to this practice, we use lowercase Greek letters to denote random error variables, for example, ε_e , as well as parameters, for example, $\beta_{e,j}$.

Moreover, assuming that the error terms in Equations 1 and 2 are independent of each other, it can be shown that the achievement index—or correlation between Y_e and Y_s —can be expressed as a multiplicative function of three terms (Tucker, 1964). These are, first, the extent to which the environment is predictable as measured by R_e , the correlation between Y_e and $\sum_{j=1}^k \beta_{e,j}X_j$; second, the consistency with which the person uses the linear decision rule as measured by R_s , the correlation between Y_s and $\sum_{j=1}^k \beta_{s,j}X_j$; and third, the correlation between the predictions of both models, that is, between $\sum_{j=1}^k \beta_{e,j}X_j$ and $\sum_{j=1}^k \beta_{s,j}X_j$. This is also known as G , the *matching index*. (Note that $G = 1$ if $\beta_{e,j} = \beta_{s,j}$, for all $j = 1, \dots, k$.)

This leads to the well-known lens model equation (Tucker, 1964) that expresses judgmental performance or achievement in the form

$$\rho_{Y_e Y_s} = GR_e R_s + \rho_{\epsilon_e \epsilon_s} \sqrt{(1 - R_e^2)(1 - R_s^2)}, \tag{3}$$

where, for completeness, we show the effect of possible nonzero correlation between the error terms of Equations 1 and 2.

Assuming that the correlation $\rho_{\epsilon_e \epsilon_s}$ is zero, we consider below two measures of judgmental performance. One is the traditional measure of achievement, $GR_e R_s$. The other is independent of the level of predictability of the environment and is captured by GR_s . Lindell (1976) referred to this as *performance*. However, we call it *linear cognitive ability*, or *ca*, to capture the notion that it measures how well someone is using the linear model in terms of both matching weights (G) and consistency of execution (R_s).

First, however, we develop the probabilities that our models make correct predictions within a given population or environment. As will be seen, these probabilities reflect the covariance structure of the cues as well as those between the criterion and the cues. It is these covariances that characterize the inferential environment in which judgments are made.

The SV Model

Imagine that the judge does not use a linear combination rule but instead simply chooses the candidate who is better on one variable, X_1 (e.g., years of experience). Thus, the decision rule is to choose the candidate for whom X_1 is larger, for example, choose A if $x_{1a} > x_{1b}$. Our question now becomes, what is the probability that A is better than B using this decision rule in a given environment, that is, what is $P\{(Y_{ea} > Y_{eb}) \cap (X_{1a} > X_{1b})\}$?

To calculate this probability, we follow the model presented in Hogarth and Karelaia (2005a). We first assume that Y_e and X_1 are both standardized normal variables (i.e., with means of 0 and variances of 1) and that the cue used is positively correlated with the criterion.⁷ Denote the correlation by the parameter $\rho_{Y_e X_1}$ (>0). Given these facts, it is possible to represent Y_{ea} and Y_{eb} by the equations

$$Y_{ea} = \rho_{Y_e X_1} X_{1a} + v_{ea} \tag{4}$$

and

$$Y_{eb} = \rho_{Y_e X_1} X_{1b} + v_{eb}, \tag{5}$$

where v_{ea} and v_{eb} are normally distributed error terms, each with mean of 0 and variance of $(1 - \rho_{Y_e X_1}^2)$, independent of each other and of X_{1a} and X_{1b} .

The question of determining $P\{(Y_{ea} > Y_{eb}) \cap (X_{1a} > X_{1b})\}$ can be reframed as determining $P\{(d_1 > 0) \cap (d_2 > 0)\}$ where $d_1 = Y_{ea} - Y_{eb} > 0$ and $d_2 = X_{1a} - X_{1b} > 0$. The variables d_1 and d_2 are bivariate normal with variance–covariance matrix

$$M_{f_{LSV}} = \begin{pmatrix} 2 & 2\rho_{Y_e X_1} \\ 2\rho_{Y_e X_1} & 2 \end{pmatrix}$$

and means of 0. Thus, the probability of correctly selecting A over B can be written as

$$\int_0^\infty \int_0^\infty f(\underline{d}) \, d\underline{d}, \tag{6}$$

where $f(\underline{d})$ is the normal bivariate probability density function with $\underline{d}' = (d_1, d_2)$.

To calculate the expected accuracy of the SV model in a given environment, it is necessary to consider the cases where both $X_{1a} > X_{1b}$ and $X_{1b} > X_{1a}$ such that the overall probability is given by $P\{((Y_{ea} > Y_{eb}) \cap (X_{1a} > X_{1b})) \cup ((Y_{eb} > Y_{ea}) \cap (X_{1b} > X_{1a}))\}$, which, because both its components are equal, can be simplified as

$$2P\{(Y_{ea} > Y_{eb}) \cap (X_{1a} > X_{1b})\} = 2 \int_0^\infty \int_0^\infty f(\underline{d}) \, d\underline{d}. \tag{7}$$

The analogous expressions for the LC, EW, CONF, and TTB models are presented in Appendix A, where the appropriate correlations for LC and EW are $\rho_{Y_e Y_s}$ and $\rho_{Y_e \bar{X}_1}$, respectively.

Loss Functions

Equation 7, as well as its analogues in Appendix A, can be used to estimate the probabilities that the models will make the correct decisions. These probabilities can be thought of as the average percentage of correct scores that the models can be expected to achieve in choosing between two alternatives. As such, this measure is equivalent to a 0/1 loss function that does not distinguish between small and large errors. We therefore introduce the notion that losses from errors reflect the degree to which predictions are incorrect.

Specifically, to calculate the expected loss resulting from using SV across a given population, we need to consider the possible losses that can occur when the model does not select the best alternative. We model loss by a symmetric squared error loss function but allow this to vary in exactingness, or the extent to which the environment does or does not punish errors severely (Hogarth, Gibbs, McKenzie, & Marquis, 1991). We note that loss occurs when (a) $X_{1a} > X_{1b}$ but $Y_{ea} < Y_{eb}$ and (b) $X_{1a} < X_{1b}$ but $Y_{ea} > Y_{eb}$. Capitalizing on symmetry, the expected loss (EL) associated with the population can therefore be written as

⁷ We consider the implications of our normality assumption in the General Discussion.

$$EL_{SV} = 2P\{(Y_{ea} < Y_{eb}) \cap (X_{1a} > X_{1b})\}L, \tag{8}$$

where $L = \alpha(Y_{eb} - Y_{ea})^2$. The constant of proportionality, $\alpha (>0)$, is the exactingness parameter that captures how heavily losses should be counted.

Substituting $\alpha(Y_{eb} - Y_{ea})^2$ for L and following the same rationale as when developing the expression for accuracy, the expected loss of the SV model can be expressed as

$$EL_{SV} = 2\alpha(Y_{eb} - Y_{ea})^2P\{(Y_{ea} < Y_{eb}) \cap (X_{1a} > X_{1b})\} \\ = 2\alpha \int_{-\infty}^0 \int_0^{\infty} d_1^2 f(d) dd. \tag{9}$$

As in the expression for accuracy, the function $f(d)$ for SV involves the variance–covariance matrix $M_{f_{-SV}}$. The expected losses of LC and EW are found analogically, using their appropriate variance–covariance matrices.

In Table 2, we summarize the expressions for accuracy and loss for SV, LC, and EW. In Appendix B, we present the formulas for the loss functions of CONF and TTB. Finally, note that expected loss, as expressed by Equation 9, is proportional to the exactingness parameter, α , that models the extent to which particular environments punish errors.

Exploring Effects of Different Environments

We first construct and simulate several task environments and demonstrate how our theoretical analyses can be used to compare

Table 2
Key Formulas for Three Models: SV, LC, and EW

Model	Variance–covariance matrix (M_f)
Single variable (SV)	$\begin{pmatrix} 2 & 2\rho_{Y,X_1} \\ 2\rho_{Y,X_1} & 2 \end{pmatrix}$
Linear combination (LC)	$\begin{pmatrix} 2 & 2\rho_{Y,Y_s} \\ 2\rho_{Y,Y_s} & 2 \end{pmatrix}$
Equal weights (EW)	$\begin{pmatrix} 2 & 2\rho_{Y,\bar{X}}\sigma_{\bar{X}} \\ 2\rho_{Y,\bar{X}}\sigma_{\bar{X}} & 2\sigma_{\bar{X}}^2 \end{pmatrix}$

Note. 1. The expected accuracy of models is estimated as the probability of correctly selecting A over B or B over A and is found as

$$2 \int_0^{\infty} \int_0^{\infty} f(d) dd,$$

where $f(d)$ is the normal bivariate probability density function with $d' = (d_1, d_2)$.

2. The expected loss of models is found as

$$2\alpha \int_{-\infty}^0 \int_0^{\infty} d_1^2 f(d) dd,$$

where $\alpha (> 0)$ is the exactingness parameter.

3. The variance–covariance matrix M_f is specific for each model.

$$4. \rho_{Y,\bar{X}} = \bar{\rho}_{Y,X} \sqrt{\frac{k}{1+(k-1)\bar{\rho}_{X,X_j}}}$$

where $k =$ number of X variables, $\bar{\rho}_{Y,X} =$ average correlation between Y and the X s, and $\bar{\rho}_{X_i,X_j} =$ average intercorrelations amongst the X s.

$$5. \sigma_{\bar{X}} = \sqrt{\frac{1}{k}(1+(k-1)\bar{\rho}_{X,X_j})}.$$

the performance of the models in terms of both expected percentage correct and expected losses. We also show how errors in the application of both linear models and heuristics affect performance and thus illustrate potential trade-offs involved in using different models. We further note that, in many environments, heuristic models achieve similar levels of performance and explicitly explore this issue using simulation. To link theory with empirical phenomena, we use a meta-analysis of lens model studies to compare the judgmental performance of heuristics with that of people assumed to be using LC models.

Constructed and Simulated Environments: Methodology

To demonstrate our approach, we constructed several sets of different three-cue environments using the model implicit in Equation 1. Our approach was to vary systematically two factors: first, the weights given to the variables as captured by the distribution of cue validities, and second, the level of average intercue correlation. As a consequence, we obtained environments with different levels of predictability, as indicated by R_e (from low to high). We could not, of course, vary these factors in an orthogonal design (because of mathematical restrictions) and hence used several different sets of designs.

For each of these, it is straightforward to calculate expected correct predictions and losses for all our models⁸ (see equations above), with one exception. This is the LC model, which requires specification of ρ_{Y,Y_s} , that is, the *achievement index*, or the correlation between the criterion and the person’s responses (see Appendix A). However, given the lens model equation—see Equation 3 above—we know that

$$\rho_{Y,Y_s} = GR_e R_s \tag{10}$$

where R_e is the predictability of the environment and GR_s or ca is the measure of linear cognitive ability that captures how well someone is using the linear model in terms of both matching weights and consistency of execution.⁹ In short, our strategy is to vary ca and observe how well the LC model performs. In other words, how accurate would people be in binary choice when modeled as if using LC with differing levels of knowledge (matching of weights) and consistency in execution of their knowledge?

For example, it is of psychological interest to ask when the validity of SV equals that of an LC strategy, that is, when $\rho_{Y,X_1} = caR_e$ or $ca = \rho_{Y,X_1}/R_e$. This is the point of indifference between making a judgment based on all the data (i.e., with LC) and relying on a single cue (SV), such as when using availability (Tversky & Kahneman, 1973) or affect (Slovic et al., 2002).

Relative Model Performance: Expected Percentage Correct and Expected Losses

We start by a systematic analysis of model performance in three sets of environments—A, B, and C—defined in Table 3. As noted

⁸ For the TTB model, we defined a threshold of .50 (with standardized variables) to decide whether a variable discriminated between two alternatives. Whereas the choice of .50 was subjective, investigation shows quite similar results if this threshold is varied between .25 and .75. We use the threshold of .50 in all further calculations and illustrations.

⁹ The assumption made here is that $\rho_{\epsilon,\epsilon_s} = 0$; see Equation 3. Recall also that *using* is employed here in an *as if* manner.

Table 3
Environmental Parameters: Cases A, B, and C

Case	Cue validities			Cue intercorrelations			R	$\rho_{Y_e X_i}/R$
	$\rho_{Y_e X_1}$	$\rho_{Y_e X_2}$	$\rho_{Y_e X_3}$	$\rho_{X_1 X_2}$	$\rho_{X_1 X_3}$	$\rho_{X_2 X_3}$		
A1	.5	.5	.5	.1	.1	.1	.81	.62
A2	.5	.5	.5	.5	.5	.5	.66	.76
B1	.7	.4	.2	.1	.1	.1	.80	.88
B2	.7	.4	.2	.5	.5	.5	.76	.93
C1	.6	.4	.3	.1	.1	.1	.75	.80
C2	.6	.4	.3	.5	.5	.5	.66	.91
C3	.6	.4	.3	-.4	.1	.1	.94	.64

above, we consider two main factors. First are distributions of cue validities. We distinguish three types: noncompensatory, compensatory, and equal weighting. Environments are classified as noncompensatory if, when cue validities are ordered in magnitude, the validity of each cue is greater than or equal to the sum of those smaller than it (cf. Martignon & Hoffrage, 1999, 2002). All other environments are compensatory. However, we distinguish between compensatory environments that do or do not involve equal weighting, treating the former as a special case. Case A in Table 3 involves equal weighting, whereas Case B is noncompensatory and Case C is compensatory (but not with equal weighting).

Second, we use average intercue correlation to define redundancy. When positive, this can be large (.50) or small (.10). It can also be negative. Thus, the variants of all cases with indices 1 (i.e., A1, B1, C1) have small positive levels of redundancy, the variants with indices 2 (i.e., A2, B2, C2) have large positive levels, and the last variant of Case C (i.e., C3) involves a negative intercue correlation.¹⁰

Taken together, these parameters imply different levels of environmental predictability (or lack of noise), that is, R_e , which varies from .66 to .94. In the right-hand column, we show values of $\rho_{Y_e X_i}/R_e$. These indicate the benchmarks for determining when SV or LC performs better. Specifically, LC performs better than SV when ca exceeds $\rho_{Y_e X_i}/R_e$.

Figure 2 depicts expected percentages of correct predictions for the different models as a function of linear cognitive ability ca . We emphasize that our models' predictions are theoretical implications as opposed to estimates of predictability gained from fitting models to data and forecasting to new samples of data. These two uses of *prediction* are quite different, and we return to this issue in the General Discussion, below.

We show only the upper part of the scale of expected percentage correct because choosing at random would lead to a correct decision in 50% of choices. We stress that, in these figures, we report the performance of SV and TTB, assuming that the cues were ordered correctly before the models were applied. We relax this assumption further below to show the effect of human error in the use of heuristics.

A first comment is that relative model performance varies by environments. In Case A1 (equal weighting and low redundancy), EW performs best (as it must). CONF is more accurate than TTB, and SV lags behind. In Case A2, where the redundancy becomes larger, the performance of all models except SV deteriorates. This is not surprising given that the increase in cue redundancy reduces

the relative validity of information provided by each cue following the first one, and thus the overall predictability of the environment decreases (i.e., R_e in Case A1 equals .81, whereas, in A2, it decreases to .66). EW, of course, still performs best. However, the other heuristic models, in particular CONF and TTB, do not lag much behind.

This picture changes in the noncompensatory environment B. When redundancy is low (i.e., Case B1), TTB performs best, followed by SV and the other heuristics. When redundancy is greater (i.e., Case B2), the performance of TTB drops some 5%. This is enough for SV, which is insensitive to the changes in redundancy, to have the largest expected performance. EW and CONF lose in performance and remain the worst heuristic performers here.

The compensatory environment C shows different trends. With low positive redundancy (i.e., Case C1), EW and TTB share the best performance, and SV is the worst of the heuristics. Higher positive cue redundancy in Case C2 allows SV to become one of the best models, sharing this position with TTB. Finally, in the presence of negative intercue correlation (i.e., Case C3), EW does best, whereas TTB stays slightly behind it. SV is again the worst heuristic. Given the same cue validities across the C environments, negative intercue correlation increases the predictability of the environment to .94 (from .75 in C1 and .66 in C2). This change triggers improvements in the performance of all models and magnifies the differences between them (compare Case C3 with Cases C1 and C2).

Now consider the performance of LC as a function of ca . First, note that, in each environment, we illustrate (by dotted vertical lines) the level of ca at which LC starts to outperform the worst heuristic. When the latter is SV, this point corresponds to the critical point of equality between LC and SV enumerated in the last column of Table 3. Thus, LC needs ca of from .62 to .80 (at least) to be competitive with the worst heuristic in these environments. The lowest demand is posed on LC in Cases A1 (minimum ca of .62) and C3 (.64). These cases are the most predictable of all examined in Figure 2 (R_e s of .81 and .94, respectively). In the least predictable environments, A2, C1, and C2, the minimum ca

¹⁰ Defining redundancy by average cue intercorrelation could, of course, be misleading by hiding dispersion among correlations. In fact, with the exception of Case C3, the intercorrelations of the variables within cases were equal—see Table 3.

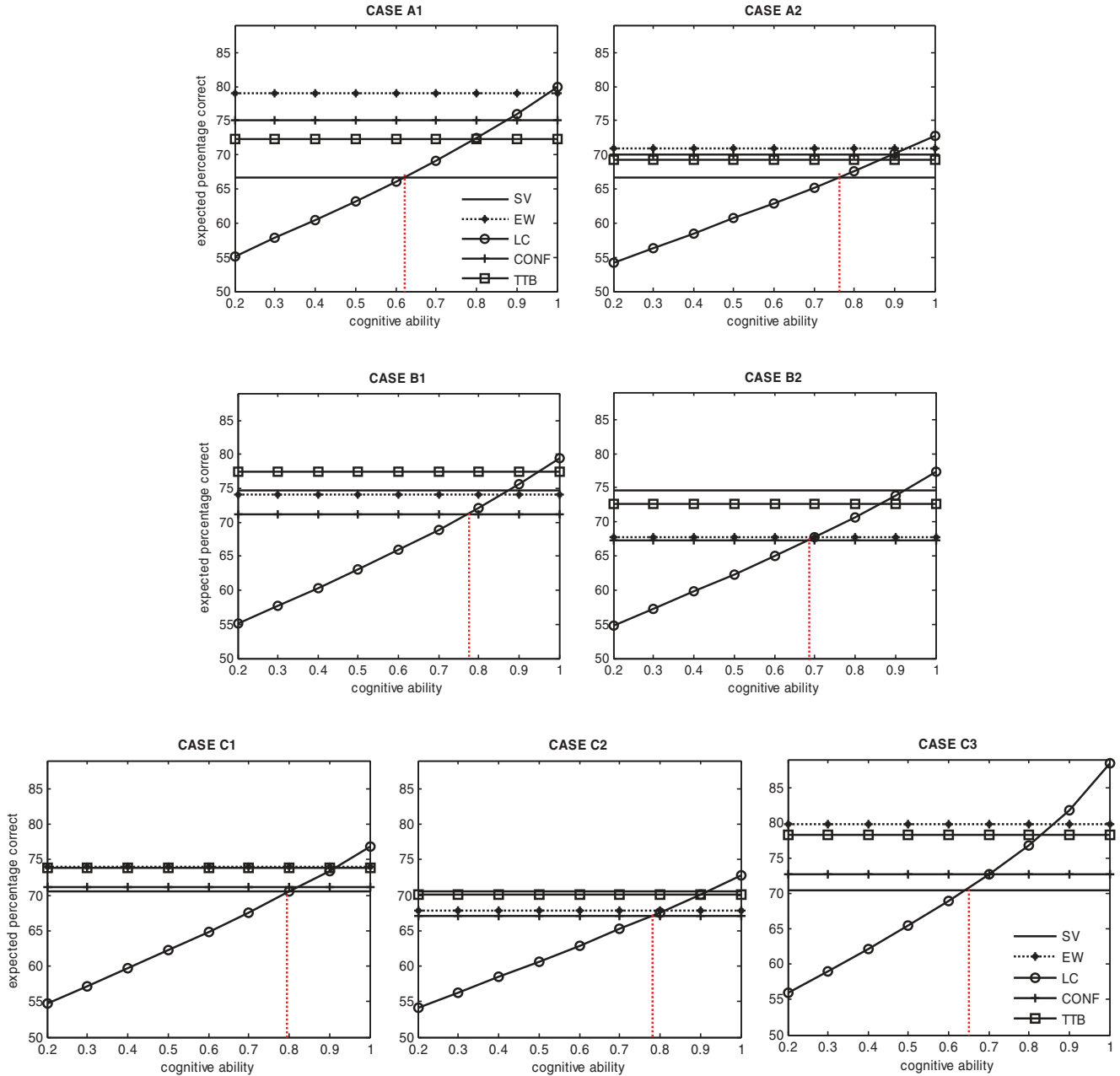


Figure 2. Model performance measured as expected percentage correct: Cases A, B, and C. SV = single variable model; EW = equal weight model; LC = linear combination model; CONF = CONF model; TTB = take-the-best model.

needed to beat the worst heuristic is much larger: .76, .80, and .78, respectively.

Interestingly, in all the environments illustrated, *ca* has to be quite high before LC starts to be competitive with the better heuristics. In the most predictable environment, C3, LC has the best performance when *ca* starts to exceed .85. In the other environments, LC starts to have the best performance only when levels of *ca* are even higher.

The simple conclusion from this analysis—which we explore further below—is that unless *ca* is high, decision makers are better

off using simple heuristics, provided that they are able to implement these correctly.

In Figure 3, we use the environment A1 to show differential performance in terms of expected loss where the exactingness parameter, α , is equal to 1.00 or .30. A comparison of expected loss with $\alpha = 1.00$ on the left panel of Figure 3 and expected accuracy in Case A1 in Figure 2 shows the same visual pattern of results in terms of relative model performance, a finding that was not obvious to us a priori. However, the differences between the models are magnified when the criterion of expected loss is used.

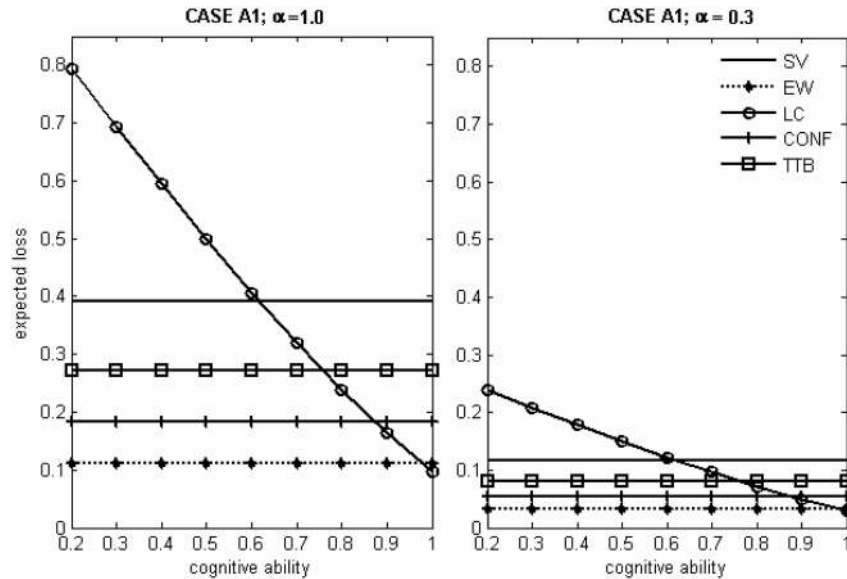


Figure 3. Model performance measured as expected loss (for $\alpha = 1.00$ and $\alpha = .30$): Case A1. SV = single variable model; EW = equal weight model; LC = linear combination model; CONF = CONF model; TTB = take-the-best model.

To note this, compare the ranges of model performance at $ca = .20$ (extreme left point) in the two figures. When expected percentage correct is used as the decision criterion, the best model (EW in this case) is some .25 points (= 80% - 55%) above the worst model (LC). When expected loss is used, the difference increases to about .70 points (= .80 of LC - .11 of SV).

The panel on the right of Figure 3 shows the effects of less exacting losses when $\alpha = .30$. Comparing it with the left panel of Figure 3, we find the same relative ordering between models but differences in expected loss are much smaller (as follows from Equation 9).

The Effect of Human Error in Heuristics

In Environments A, B, and C, we assume that the cues are ordered correctly before the heuristics are applied. However, this excludes the possibility of human error in executing the heuristics. To provide more insight, we relax this assumption in a further set of environments D. Similar to the environments described above, we consider two variants of D: D1, with low positive cue redundancy, and D2, with a higher level of redundancy (see Table 4). To show additionally the effect of predictability, R_e , within environments, we include eight subcases (i-viii) in both variants. The distribution of cue validities is noncompensatory in Subcases i, ii, and iii; compensatory in Subcases iv, v, and vi; and equal weighting in the last two subcases, vii and viii. A consequence of these specifications is a range of environmental predictabilities, R_e , from .37/.39 to .85/.88 across all eight sets of subcases.

In Table 4, we report both expected percentage correct and losses (for $\alpha = 1.00$) for all models. To illustrate effects of human error, we present heuristic performance under the assumption that the decision maker fails to order the cues according to their validities and thus uses them in random order. This error affects the results of SV and TTB. EW and CONF, however, are immune

to this lack of knowledge of the environmental structure. For SV and TTB, we present in addition results achieved with correct knowledge about cue ordering. To illustrate the effect of knowledge on the performance of LC in the same environments, we show results using three values for ca : $ca = .50$ for LC1, $ca = .70$ for LC2, and $ca = .90$ for LC3.

The trends in Table 4 are illustrated in Figures 4 and 5, which document percentage correct and expected loss, respectively, of the different models as a function of the validity of the most valid cue, ρ_{Y,X_1} . Because, here, ρ_{Y,X_1} is highly correlated with environmental predictability R_e , the horizontal axis of the graphs can also be thought of as capturing noise (more, on the left, to less, on the right).

In the upper (lower) panel of the figures, we show the effect of error on the performance of SV (TTB). The performances of SV and TTB under random cue ordering are illustrated with the corresponding lines. The range of possible performance levels of the models from best (i.e., achieved under the correct cue ordering) to worst (i.e., achieved when the least valid cue is examined first, the second least valid second, etc.) is illustrated with the shaded areas.

First, we compare performance among the heuristic models. Note that, as noise in the environment decreases, there is a general trend for differences in heuristic model performance to increase, in addition to a tendency for performance to improve (see Figure 4). Second, relative model performance is affected by distributions of cue validities and redundancy (see Table 4). In noncompensatory environments with low redundancy (Subcases i-iii), TTB performs best, provided that the cues are ordered correctly (Figure 4, lower panel, the right-hand part of Case D1, the upper limit of the range of TTB). However, as these environments become more redundant, the advantage goes to SV (Figure 4, upper panel, the right-hand side of Case D2, the upper limit of the range of SV). When

Table 4
The Effect of Human Error on Model Performance: Case D

Case & subcase	Environmental parameters										Loss ($\alpha = 1.00$)																		
											Percentage correct						SV												
	Cue validities			Cue redundancy			ρ_{Y_X}/R_e				SV		TTB		SV		TTB		SV		TTB								
	1	2	3	1	2	3	LC1	LC2	LC3	EW	CONF	LC1	LC2	LC3	EW	CONF	LC1	LC2	LC3	EW	CONF	LC1	LC2	LC3	EW	CONF			
D1																													
i	0.8	0.4	0.2	0.1			0.91	64	71	79	80	66	82	72	76	74	74	0.5	0.3	0.1	0.1	0.5	0.1	0.1	0.5	0.1	0.3	0.2	0.2
ii	0.7	0.4	0.2				0.88	63	69	76	75	65	78	69	74	71	71	0.5	0.3	0.2	0.2	0.5	0.2	0.2	0.5	0.1	0.4	0.2	0.3
iii	0.6	0.4	0.2				0.73	62	67	73	70	63	73	67	72	69	69	0.6	0.4	0.2	0.3	0.5	0.2	0.3	0.5	0.2	0.4	0.3	0.3
iv	0.5	0.4	0.2				0.66	61	65	70	67	62	70	66	70	67	67	0.6	0.4	0.3	0.4	0.5	0.3	0.4	0.5	0.3	0.4	0.3	0.4
v	0.4	0.4	0.2				0.61	60	64	69	63	61	66	64	68	66	66	0.6	0.5	0.3	0.5	0.6	0.4	0.5	0.6	0.4	0.5	0.4	0.4
vi	0.3	0.3	0.2				0.52	58	62	66	60	59	62	61	64	62	62	0.7	0.5	0.4	0.6	0.7	0.6	0.6	0.7	0.6	0.6	0.5	0.5
vii	0.2	0.2	0.2				0.45	57	60	63	56	56	58	58	60	59	59	0.7	0.6	0.5	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.6	0.6
viii	0.1	0.1	0.1				0.39	56	59	61	53	53	54	54	55	55	55	0.8	0.7	0.6	0.9	0.9	0.8	0.9	0.8	0.8	0.8	0.8	
D2																													
i	0.8	0.4	0.2	0.5			0.85	64	70	78	80	66	76	68	69	69	69	0.5	0.3	0.1	0.1	0.5	0.2	0.2	0.5	0.2	0.4	0.3	0.4
ii	0.7	0.4	0.2				0.76	62	68	74	75	65	73	67	68	67	68	0.5	0.4	0.2	0.2	0.5	0.2	0.2	0.5	0.2	0.4	0.4	0.4
iii	0.6	0.4	0.2				0.67	61	66	71	70	63	70	65	66	66	66	0.6	0.4	0.3	0.4	0.5	0.3	0.5	0.3	0.5	0.4	0.4	0.4
iv	0.5	0.4	0.2				0.60	60	64	68	67	62	67	64	65	64	65	0.6	0.5	0.3	0.4	0.5	0.4	0.5	0.5	0.5	0.5	0.5	0.5
v	0.4	0.4	0.2				0.54	59	62	66	63	61	64	62	63	63	63	0.7	0.5	0.4	0.5	0.6	0.5	0.6	0.6	0.5	0.5	0.5	0.5
vi	0.3	0.3	0.2				0.47	58	61	64	60	59	61	60	61	60	60	0.7	0.6	0.5	0.6	0.7	0.6	0.6	0.7	0.6	0.6	0.6	0.6
vii	0.2	0.2	0.2				0.41	57	59	62	56	56	57	57	58	58	58	0.7	0.6	0.5	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
viii	0.1	0.1	0.1				0.37	56	58	61	53	53	54	54	54	54	54	0.8	0.7	0.6	0.9	0.9	0.9	0.9	0.9	0.9	0.8	0.9	

Note. For LC1, $ca = 0.5$; for LC2, $ca = 0.7$; for LC3, $ca = 0.9$. The performance of the best heuristic in each environment is highlighted with boldface characters. The performance of LC is underlined and presented on a darker background when it is superior or equal to that of the best performer among heuristics. LC = linear combination model; SV = single variable model; TTB = take-the-best model; EW = equal weight model; CONF = CONF model; ca = linear cognitive ability.

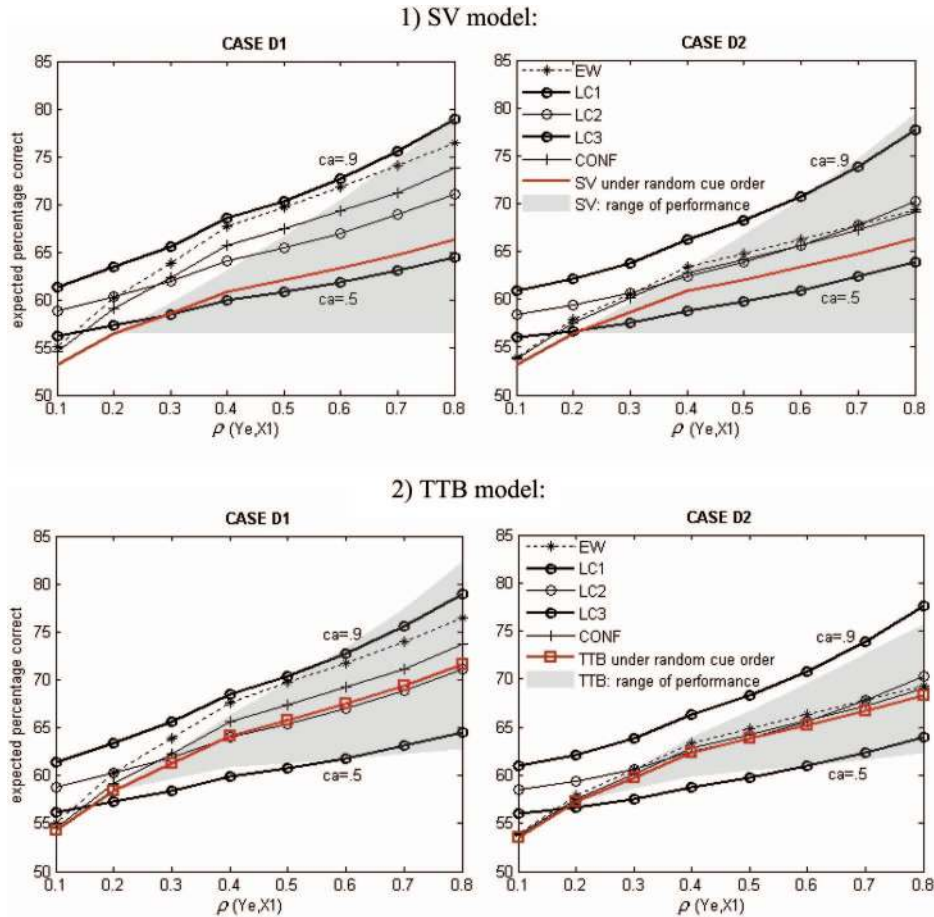


Figure 4. The effect of human errors on expected percentage correct: Case D. SV = single variable model; EW = equal weight model; LC = linear combination model; CONF = CONF model; TTb = take-the-best model.

environments involve equal weighting (Subcases vii and viii), EW is the most accurate, followed by CONF. In the compensatory environments (Subcases iv–vi), EW does best when redundancy is low, but this advantage switches to TTb (provided that the cues are ordered correctly) when redundancy is higher. We discuss these trends again below.

Second, comparing Figures 4 and 5, we note again that expected loss rank-orders the models similarly to expected percentage correct. The differences among the models are more pronounced and evident, however, when expected loss is used.

Third, extreme errors in ordering the cues according to their validities decrease the performance of SV and TTb so much that even in the most predictable environments (observe the lower bounds of SV and TTb at the right-hand side of illustrations in Figures 4 and 5), this can fall almost to the levels of performance corresponding to the most noisy environments (same bounds at the left-hand side of the illustrations). In addition, SV is punished relatively more than TTb by ordering the cues incorrectly (compare the vertical widths of the SV and TTb shaded ranges in both Figures 4 and 5). When knowledge about the structure of the environment is lacking, more extensive cue processing under EW and CONF hedges the decision maker irrespective of the type of environment (i.e., compensatory or not).

Note that, in equal-weighting environments (i.e., Subcases vii and viii), $\rho_{Y_e, X_1} = .10$ and $.20$, it does not matter whether SV and TTb identify the correct ordering of cues because each has the same validity. In these environments, the ranges of performance of SV and TTb coincide with the model performance under random cue ordering.

Fourth, when expected loss is used instead of expected percentage correct, the decrease in performance due to incorrect cue ordering is more pronounced. This is true for both SV and TTb. (Compare the vertical width of the shaded ranges between Figures 4 and 5, within the models. Note that the scales used in Figures 4 and 5 are different and that using equivalent scales would mean decreasing all the vertical differences in Figure 4).

Fifth, for the LC model, it is clear (and unsurprising) that more ca is better than less. Interestingly, as the environment becomes more predictable, the accuracy of the LC models drops off relative to the simpler heuristics. In the environments examined here, the best LC model (with $ca = .90$) is always outperformed by one of the other heuristics when $\rho_{Y_e, X_1} > .60$ (see Table 4). Error in the application of heuristics, however, can swing the advantage back to LC models even in the most predictable environments (the right-hand side of illustrations in Figure 4, below the upper bounds of SV and TTb). In addition, errors in the application of heuristic

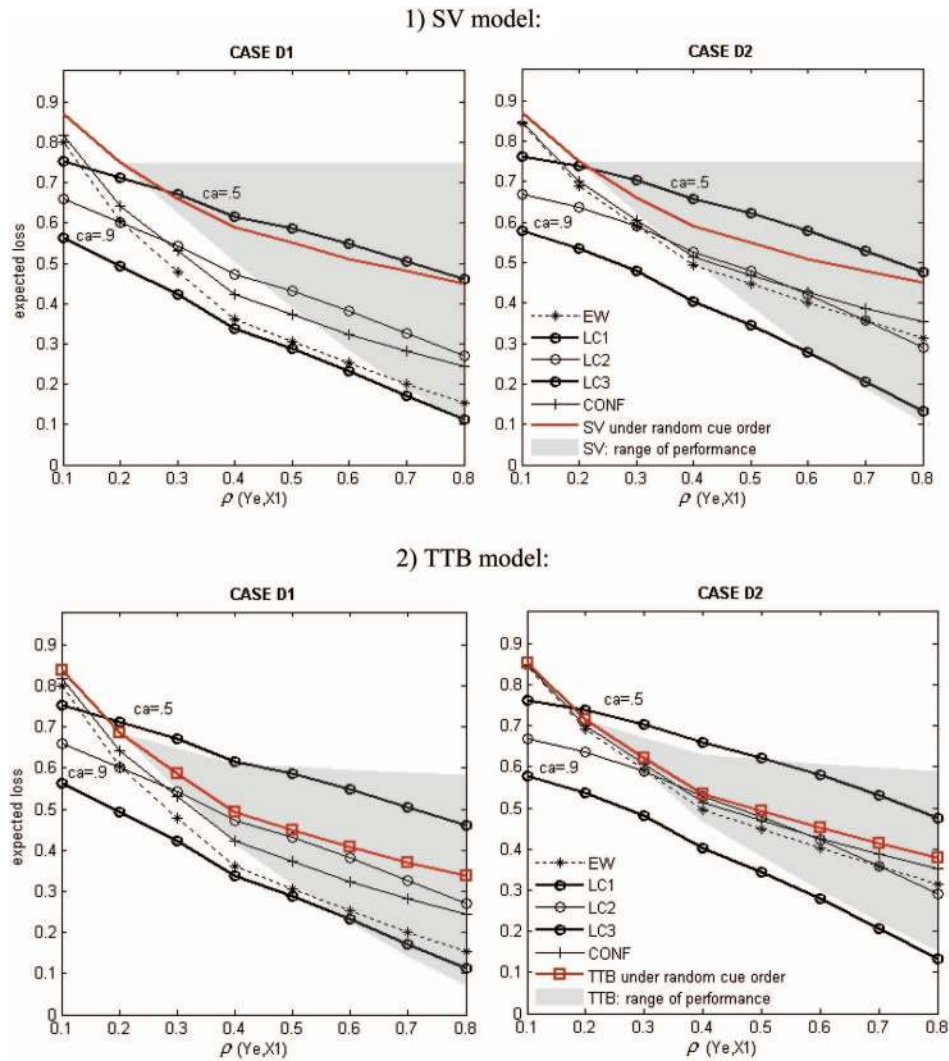


Figure 5. The effect of human errors on expected loss (for $\alpha = 1.00$): Case D. SV = single variable model; EW = equal weight model; LC = linear combination model; CONF = CONF model; TTB = take-the-best model.

models mean that LC can be relatively more accurate at the lower levels of ca .

Agreement Between Models

In many instances, strategies other than LC have quite similar performance. This raises the question of knowing how often they make identical predictions. To assess this, we calculated the probability that all pairs of strategies formed by SV, EW, TTB, and CONF would make the same choices across several environments. In fact, because calculating this joint probability is complicated, we simulated results on the basis of 5,000 trials for each environment.

Table 5 specifies the parameters of the E environments, the percentage of correct predictions for each model in each environment,¹¹ and the probabilities that models would make the same decisions. There are two variants, E1 and E2 (with low and higher

redundancy), each with eight subcases (i–viii). For both cases, the environments of Subcases i–v are noncompensatory, Subcase vi is compensatory, and Subcases vii and viii involve equal weighting. Across each case, predictability (R_e) varies from high to low.

We make three remarks. First, there is considerable variation in percentage-correct predictions across different levels of predictability that are consistent with the results reported in Table 4. However, agreement between pairs of models hardly varies as a function of R_e and is uniformly high. In particular, the rate of agreement lies between .70 and .92 across all comparisons and is

¹¹ We also calculated the theoretical probabilities of the simulated percentage of correct predictions. Given the large sample sizes (5,000), theoretical and simulated results are almost identical.

Table 5
Rates of Agreement Between Heuristic Strategies for Different Environments

Case & subcase	Cue validities			Cue redundancy	R_e	Percentage correct				Rates of agreement					
	1	2	3			SV	EW	TTB	CONF	SV-EW	SV-CONF	SV-TTB	CONF-EW	TTB-EW	CONF-TTB
E1															
				0.1											
i	0.8	0.6	0.2		0.96	80	82	86	79	0.72	0.77	0.72	0.87	0.80	0.77
ii	0.7	0.5	0.2		0.84	75	77	79	74	0.73	0.77	0.73	0.86	0.80	0.77
iii	0.6	0.4	0.2		0.73	71	72	73	69	0.72	0.77	0.72	0.86	0.80	0.77
iv	0.5	0.3	0.2		0.63	66	67	67	66	0.71	0.76	0.71	0.85	0.78	0.76
v	0.4	0.2	0.2		0.54	62	63	63	61	0.73	0.77	0.73	0.86	0.80	0.78
vi	0.3	0.2	0.2		0.49	59	61	61	61	0.70	0.76	0.70	0.85	0.78	0.76
vii	0.2	0.2	0.2		0.45	57	59	58	58	0.72	0.78	0.72	0.86	0.79	0.76
viii	0.1	0.1	0.1		0.38	53	54	53	53	0.71	0.77	0.71	0.85	0.79	0.76
					Mean	<u>65</u>	<u>67</u>	<u>68</u>	<u>65</u>	<u>0.72</u>	<u>0.77</u>	<u>0.72</u>	<u>0.86</u>	<u>0.79</u>	<u>0.77</u>
E2															
				0.5											
i	0.8	0.6	0.2		0.90	80	73	79	71	0.81	0.83	0.81	0.91	0.88	0.85
ii	0.7	0.5	0.2		0.78	74	70	74	68	0.81	0.84	0.81	0.91	0.88	0.85
iii	0.6	0.4	0.2		0.67	71	67	70	65	0.81	0.83	0.81	0.91	0.88	0.84
iv	0.5	0.3	0.2		0.58	67	64	66	63	0.80	0.83	0.80	0.91	0.88	0.85
v	0.4	0.2	0.2		0.50	64	61	63	60	0.80	0.83	0.80	0.91	0.87	0.84
vi	0.3	0.2	0.2		0.44	59	59	60	59	0.81	0.84	0.81	0.91	0.87	0.84
vii	0.2	0.2	0.2		0.42	58	59	58	59	0.80	0.83	0.80	0.91	0.88	0.85
viii	0.1	0.1	0.1		0.38	53	54	53	53	0.80	0.83	0.80	0.92	0.88	0.84
					Mean	<u>66</u>	<u>63</u>	<u>65</u>	<u>62</u>	<u>0.81</u>	<u>0.83</u>	<u>0.81</u>	<u>0.91</u>	<u>0.88</u>	<u>0.85</u>
					Overall mean	<u>66</u>	<u>65</u>	<u>66</u>	<u>64</u>	<u>0.76</u>	<u>0.80</u>	<u>0.76</u>	<u>0.88</u>	<u>0.84</u>	<u>0.81</u>

Note. Results are from simulations with 5,000 trials for each environment. SV = single variable model; EW = equal weight model; TTB = take-the-best model; CONF = CONF model.

probably higher than one might have imagined a priori.¹² At the same time, this means that differences between the models occur in 8%–30% of choices, and from a practical perspective, it is important to know when this happens and which model is more likely to be correct. Second, as would be expected, the effect of increasing redundancy is to increase the level of agreement between models. Third, for the environments illustrated here, CONF and EW have the highest level of agreement whereas SV-EW and SV-TTB have the lowest. The latter result is surprising in that both SV and TTB are so dependent on the most valid cue.

Relative Model Performance: A Summary

Synthesizing the results of the 39 environments specified in Tables 3, 4, and 5, we can identify several trends in the relative performance of the models.

First, the models all perform better as the environment becomes more predictable. At the same time, differences in model performance grow larger.

Second, relative model performance depends on both how the environment weights cues (noncompensatory, compensatory, or equal weighting) and redundancy. We find that when cues are ordered correctly, (a) TTB performs best in noncompensatory environments when redundancy is low; (b) SV performs best in noncompensatory environments when redundancy is high; (c) irrespective of redundancy, EW performs best in equal-weighting environments in which CONF also performs well; (d) EW (and sometimes TTB) performs best in compensatory environments when redundancy is low; and (e) TTB (and sometimes SV) performs best in compensatory environments when redundancy is high.

Third, subject to the differential predictive abilities noted, the heuristic models exhibit high rates of agreement.

Fourth, any advantage of LC models falls sharply as environments become more predictable. Thus, a high level of *ca* is required to outpredict the best heuristics. On the other hand, error in the execution of heuristics can result in more accurate performance by LC models.

Fifth, when the decision maker does not know the structure of the environment and therefore cannot order the cues according to their validity, the more extensive EW and CONF models are the best heuristics, irrespective of how the environment weights cues and redundancy. This is an important result in that it justifies use of these heuristics when decision makers lack knowledge of the environment, that is, these are good heuristics for states of comparative ignorance (see also Karelaia, 2006).

We discuss this summary again below.

Comparisons With Experimental Data

The above analysis has been at a theoretical level and raises the issue of how good people are at making decisions with linear models as opposed to using heuristics. To answer this question, we undertook a meta-analysis of lens model studies to estimate *ca*. This involved attempting to locate all lens model studies reported in the literature that provided estimates of the elements of Equation

¹² In the populations A–C, the analogous rates of agreement were .64 to .92. Interestingly, it was the environment with negative intercue correlation that had the lowest rates of agreement (mean agreement between models .70).

Table 6
Description of Studies in Lens Model Meta-Analysis

Characteristics of tasks	Number of studies	Average number		Mean lens model statistics						LC accuracy (%)
		Judges	Judgments	r_a^a	G^a	R_e	R_s	C^a	GR_s	
All studies	270	20	86	0.56	0.80	0.79	0.80	0.05	0.66	70
Participants										
Experts	61	15	153	0.53	0.74	0.72	0.83	0.12	0.62	69
Novices	206	20	66	0.57	0.83	0.82	0.80	0.03	0.67	71
Unclassified	3	53	66	0.24	0.45	0.72	0.74	0.00	0.24	58
Type of study										
Laboratory	214	20	86	0.57	0.83	0.80	0.80	0.04	0.68	71
Field	53	16	87	0.50	0.72	0.75	0.83	0.08	0.60	68
Unclassified	3	22	96	0.41	0.52	0.81	0.88	0.00	0.44	65
Number of cues										
2	69	26	48	0.63	0.88	0.79	0.79	0.07	0.71	73
3	90	19	93	0.55	0.88	0.80	0.81	0.00	0.72	70
>3	108	16	105	0.52	0.71	0.79	0.81	0.07	0.58	69
Unclassified	3	21	56	0.17	0.32	0.74	0.86	-0.01	0.02	56
Type of weighting function										
Equal weighting	42	29	65	0.66	0.91	0.82	0.81	0.02	0.75	75
Compensatory	91	16	97	0.56	0.83	0.79	0.82	0.04	0.68	70
Noncompensatory	60	22	40	0.56	0.81	0.85	0.75	0.04	0.64	70
Unclassified	77	16	120	0.49	0.72	0.74	0.82	0.08	0.60	68
Cue redundancy ^b										
None	106	20	50	0.61	0.89	0.82	0.81	0.03	0.73	72
Low-medium	89	19	91	0.55	0.78	0.79	0.83	0.03	0.65	70
High	25	26	101	0.54	0.76	0.76	0.80	0.10	0.64	69
Unclassified	50	15	150	0.48	0.72	0.76	0.74	0.10	0.53	67

Note. LC = linear combination model.

^a These statistics correspond to the sample estimates of the elements of the lens model equation presented in the text—Equation 3 (r_a is the estimate of the “achievement” index, $\rho_{Y_e Y_s}$; G is the estimate of the matching index; and C is the estimate of the correlation between residuals of the models of the person and the environment, ρ_{e, e_s}).

^b We define redundancy level by the average intercue correlation: None denotes 0.0; low-medium denotes absolute value ≤ 0.4 ; and high otherwise.

3. Studies therefore had to have a criterion variable and involve the judgments of individuals (as opposed to groups of people).¹³ Moreover, we considered only cases in which there was more than one cue (with one cue, $G = 1.00$ necessarily). We located 84 (mainly) published studies that allowed us to examine judgmental performance across 270 different task environments (i.e., environments that vary by statistical parameters and/or substantive conditions).¹⁴

In Table 6, we summarize statistics from the meta-analysis (for full details, see Karelaia & Hogarth, 2007). First, we note that these studies represent much data. They are the result of approximately 5,000 participants providing a total of some 320,000 judgments. In fact, many of these studies involved learning, and because we characterize judgmental performance by that achieved in the last block of experimental trials reported, the participants actually made many more judgments. Second, we provide several breakdowns of different lens model and performance statistics that are the means across studies of individual data that have been averaged within studies (i.e., the units of analysis are the mean data of particular studies). We distinguish between expert and novice participants, laboratory and field studies, environments that involved different numbers of cues, different weighting functions, and different levels of redundancy (or cue intercorrelation).

Briefly, we find no differences in performance between participants who are experts or novices (the latter, however, are assessed

after learning). Holding the predictability of the environment constant (i.e., R_e), performance is somewhat better with fewer cues and when the environment involves equal weighting as opposed to being compensatory or noncompensatory. Controlling for the number of the cues, there is no difference in performance between laboratory and field studies.

Overall, the LC accuracy reported in the right-hand column of Table 6 is about 70%. This represents the percentage correct in binary choice of a person whose estimated linear cognitive ability (ca or GR_s) is .66. Moreover, this figure is a mean estimate across individual studies, each of which is described by the mean of individual data. Table 6 obscures individual variation, which we discuss further below.

To capture the differences in performance between LC and the heuristic models, one needs to have specific information on the statistical properties of tasks (essentially the covariation matrix

¹³ We also excluded studies from the interpersonal conflict paradigm in which the criterion for one's person's judgments is the judgment of another person (see, e.g., Hammond, Wilkins, & Todd, 1966).

¹⁴ It is important to bear in mind that, although investigators in the lens model paradigm model judgments as though people are using linear models, judges may, in fact, be using quite different processes (Michael E. Doherty, personal communication, July 2006).

Table 7
Performance of Heuristics and Mean LC in Three-Cue and Two-Cue Environments

Weighting function	Maximum possible percentage correct	Performance—percentage correct							Numbers of environments
		LC ^a	SV	SVr	EW	CONF	TTB	TTBr	
Three-cue environments ^b									
Equal weighting	81	72	65	65	80	74	71	70	9
Compensatory	79	69	69	64	76	71	73	67	25
Noncompensatory	81	68	74	64	75	71	78	68	30
								Subtotal	<u>64</u>
Two-cue environments ^c									
Equal weighting	88	77	71	71	87	71	78	78	17
Noncompensatory	84	69	76	67	73	67	75	70	21
								Subtotal	<u>38</u>
								Total	<u>102</u>

Note. Boldface indicates largest percentage correct in each row. LC = linear combination model; SV = single variable model; SVr = SV executed under random cue order; EW = equal weight model; CONF = CONF model; TTB = take-the-best model; TTBr = TTB executed under random cue order.

^a Based on empirically observed mean linear cognitive ability (*ca*).

^b Averages calculated on the 64 environments detailed in Appendix C.

^c Averages calculated on the 38 environments detailed in Appendix D.

used to generate the environmental criterion) and to make predictions for each environment. Recall also that, in the lens model paradigm, performance—or achievement—is measured in terms of correlation. We therefore transformed this measure into one of performance in binary choice using the methods described above. Thus, to estimate the accuracy of LC relative to any heuristic in a particular environment, we considered the difference in expected predictive accuracies between LC based on the mean *ca* observed in the environment and that of the heuristic. In other words, we asked how well the average performance levels of humans using LC compare with those of heuristics.

In Table 7, we summarize this information for environments involving three and two cues (details are provided in Appendices C and D). Unfortunately, not all studies in our meta-analysis provided the information needed, and thus, our data are limited to approximately two thirds of tasks involving three cues and one half of tasks involving two cues. We also note, parenthetically, that although some environments had identical statistical properties, they can be considered different because they involved different treatments (e.g., how participants had been trained, different feedback conditions, presentation of information, etc.).

The upper panel of Table 7 summarizes the data from Appendix C. The first column (on the left) shows the maximum performance that could be achieved in environments characterized by equal-weighting, compensatory, and noncompensatory functions, respectively. This captures the predictability of the environments—81% for equal-weighting and noncompensatory and 79% for compensatory environments. These environments are also marked by little redundancy. About 77% have mean intercue correlations of 0.00. In the body of the table, we present performance in terms of percentage correct for LC—based on mean *ca* observed in each of the experimental studies—as well as the performance that would have been achieved by the different heuristics in those same environments.

As would be expected, the EW strategy performs best in equal-weighting environments (80%) and the TTB strategy best in the

noncompensatory environments (78%). Interestingly, in the compensatory environments here, it is the EW model that performs best (76%). The mean LC model never has the best performance. Compared with the heuristic models, its performance is relatively better in the equal-weighting as opposed to the other environments.

In the discussion so far, we have concentrated on effects of error in using LC (by focusing on *ca*). However, the columns headed SVr and TTBr illustrate the effects of making errors when using heuristics (the suffix *-r* indicating models with random cue orderings).¹⁵ This shows that the performance of LC (at mean *ca* level) is as good as or better than SVr and TTBr across all three types of environments.

In the lower panel of Table 7, we present the data based on analyzing studies with two cues, where, once again, most environments involve orthogonal cues (76%)—details are provided in Appendix D. Conclusions are similar to the three-cue case. EW is necessarily best when the environment involves an equal-weighting function, and TTB performs well in the noncompensatory environments, although it is bettered here by the SV model (just).¹⁶

Because most published studies do not report individual data, it is difficult to assess the importance of individual variation in performance and, specifically, how individual LC performance compares with heuristics. Two studies involving two cues did report the necessary data (Steinmann & Doherty, 1972; York, Doherty, & Kamouri, 1987). Table 8 summarizes the comparisons. This shows (reading from left to right) the number of participants in each task, statistical properties of the tasks, percentage performance correct by the LC model (mean and range), and the per-

¹⁵ The TTBr model is identical to what Gigerenzer et al. (1999) referred to as MINIMALIST.

¹⁶ The following rule was used to adapt the CONF model for two cues: If both cues suggest the same alternative, choose it. Otherwise, choose at random.

Table 8
Levels of Individual Performance Relative to Heuristics

Study	Number of participants	Statistical properties of tasks				LC performance (% correct)			Percentage of participants with better LC performance than:					
		R_e	$\rho_{Y_cX_1}$	$\rho_{Y_cX_2}$	$\rho_{X_1X_2}$	Mean	Maximum	Minimum	SV	SVr	TTB	TTBr	EW	CONF
Steinmann & Doherty (1972)	22	0.95	0.69	0.65	0.00	73	85	58	45	50	18	18	0	50
York, Doherty, & Kamouri (1987)														
Group 1	15	0.86	0.78	0.37	0.00	70	84	53	7	57	7	36	7	57
Group 2	15	0.86	0.78	0.37	0.00	67	78	54	0	29	0	21	0	29
Group 3	15	0.86	0.78	0.37	0.00	72	80	54	14	71	0	57	0	71

Note. LC = linear combination model; SV = single variable model; SVr = SV executed under random cue order; TTB = take-the-best model; TTBr = TTB executed under random cue order; EW = equal weight model; CONF = CONF model.

centages of participants who have better performance with LC than with particular heuristics.

Clearly, one cannot generalize from the four environments presented in Table 8. However, it is of interest to note, first, that there is a large range of individual LC performances and, second, that for a minority of participants, LC performance is better than that of heuristics.

Summary

At a theoretical level, we have shown that the performance of heuristic rules is affected by several factors: how the environment weights cues, that is, noncompensatory, compensatory, or equal weighting; cue redundancy; the predictability of the environment; and loss functions. Heuristics work better when their characteristics match those of the environment. Thus, EW predicts best in equal-weighting situations and TTB in noncompensatory environments. However, redundancy allows SV to perform better than TTB in noncompensatory environments. When environments are compensatory, redundancy further mediates the relative performances of TTB, SV, and EW (TTB and SV are better with redundancy). As environments become more predictable, all models perform better, but differences between models also increase. However, when the environmental structure is unknown, the heuristics involving more extensive information processing, EW and CONF, dominate the lexicographic-type simple models, that is, SV and TTB, irrespective of cue redundancy and of how environments weight cues. Finally, the effect of loss functions is to accentuate or dampen differences between evaluations of model predictions.

We have also used simulation to investigate the extent to which models agree with each other. At one level, all the models we investigated are sensible and use valid information. As such, they exhibit much agreement. The extent of the agreement, however, is surprising. Even when the predictability of the environment varies greatly, the level of agreement between particular models hardly changes (see Table 5). From a predictive viewpoint, this might be thought comforting. However, it also accentuates the need to know which heuristic is more likely to be correct in the 8%–30% of cases in which they disagree. In addition, whereas some differences between models may seem small on single occasions, cumulative effects could be large if people were to persist in using inappropriate heuristics across many decisions.

The differential impact of environmental factors is illustrated quantitatively in Table 9, which reports the results of regressing performance of the heuristics (percentage correct) on environmental factors: type of weighting function, redundancy (cue intercorrelation), and predictability (R_e). This is done for the 39 populations specified in Tables 3, 4, and 5. Results show the importance of noncompensatory and compensatory environments (vs. equal weighting) as well as of redundancy on SV (positive). Both EW and CONF depend (negatively) on whether environments are noncompensatory, EW being affected additionally by redundancy (negatively). Interestingly, for the conditions examined here, the performance of TTB is not affected by these factors (it is fully explained by predictability R_e), thereby suggesting a heuristic that is robust to environmental variations (as has also been demonstrated theoretically by Hogarth & Karelaia, 2005b, 2006b; and Baucells, Carrasco, & Hogarth, in press). Finally, all these models benefit from greater predictability.

When cues are ordered at random, the SV and TTB models (denoted by SVr and TTBr, respectively) become less dependent on predictability R_e (compare also the intercepts for SV and SVr and for TTB and TTBr). The LC model is explained almost equally by environmental predictability, R_e , and linear cognitive ability, ca or GR_s —see Regression LC(a). On the other hand, when purposely omitting predictability, R_e , from the LC regression, compensatory and noncompensatory characteristics (vs. equal weighting) become significant (positive), as well as redundancy (negative)—see regression LC(b). Moreover, the value of the intercept increases (from 29.4 to 48.2).

An important conclusion from our theoretical analysis is that unless ca is high, people are better off relying on trade-off-avoiding heuristics as opposed to linear models. At the same time, however, the application of heuristic rules can involve error (i.e., variables not used in the appropriate order in SV and TTB). This therefore raises the issue of estimating ca from empirical data and noting when this is large enough to do without heuristics.

Our theoretical analyses suggest that ca needs to be larger than about .70 for LC models to perform better than heuristics. Across the 270 task environments of the meta-analysis, we estimate ca to be .66. However, this is a mean and does not take account of differences in task environments. For those environments in which precise predictions could be made, LC models based on mean ca estimates perform at a level inferior to the best heuristics but equal

Table 9
Regression of Model Performance (Percentage Correct) on Environmental Characteristics for Populations in Tables 3, 4, and 5

Independent variable	SV		SVr		TTB		TTBr		EW		CONF		LC (a)		LC (b)		
	Regression coefficient	t	Regression coefficient	t	Regression coefficient	t	Regression coefficient	t	Regression coefficient	t	Regression coefficient	t	Regression coefficient	t	Regression coefficient	t	
Intercept	34.3	30.7	43.9	50.9	36.6	37.6	41.7	44.2	41.7	41.7	38.6	41.6	34.5	29.4	36.8	48.2	40.7
Dummy: compensatory	2.2*	3.1														2.3*	2.9
Dummy: noncompensatory	4.9	6.1					-1.7*	-3.0	-2.0*	-2.6	-1.5*						
Redundancy	6.7	5.4	4.1	4.3					-4.8*	-7.9						4.6	5.7
Predictability (R_e)	39.2	20.0	25.7	22.2	49.1	33.2	35.8	21.8	43.0	38.4	18.2	38.1	18.2	25.9	28.9	-7.0	-4.7
Linear cognitive ability (GR_s)																	
Adjusted R^2	0.97		0.93		0.97		0.94		0.94			0.92		0.95		24.6	17.8
																0.73	

Note. The regressions are based on 39 observations, except for the regression explaining LC, which is based on 127 observations. The dummy variables for compensatory and noncompensatory weighting functions are expressed relative to equal weighting, the effect of which is captured within the intercept term. There are only three levels of redundancy: mean intercue correlation of -0.07, 0.1, and 0.5. Only statistically significant coefficients are shown ($p < .001$; except where marked with an asterisk). SV = single variable model; SVr = SV executed under random cue order; TTB = take-the-best model; TTBr = TTB executed under random cue order; EW = equal weight model; CONF = CONF model; LC = linear combination model.
* $p < .05$.

to or better than heuristics executed with error. Unfortunately, the data do not allow us to make a thorough investigation of individual variation in *ca* values.

General Discussion

Our article has shown how different views of heuristic decision making can be reconciled within a framework that also encompasses the representation of human judgment as linear models. Central to our work is the importance of understanding the effects of different environments that we have characterized by their statistical properties. We now consider implications that are, first, psychological; second, normative; and third, methodological. We also outline extensions for further work.

Psychological Implications

All of the models (heuristics) we have examined represent ideal types. Thus, it is legitimate to ask how their mathematical representations capture underlying psychological processes. This is not a new issue (see, e.g., Einhorn et al., 1979; Hoffman, 1960). Apart from predictive tests, we believe the answer lies in assessing consistency between the assumptions of models and the information-processing operations actually performed by humans.

Consider, for example, the models that are arguably the most simple and complex, namely, the SV and LC models. For the former, we can argue that the psychological process is modeled correctly if the assumption that the judgment is based on a single cue is verified. It does not matter, for example, if the individual looks at other cues and then ignores them. For the latter, checking for consistency is more complex. Were all cues examined? Were weights attached to the cues? Were the weighted sums aggregated to form a global judgment? Note that there is no need to say that actual mathematical formulae were used. All one needs to show is that mental operations that led to outcomes consistent with the operations took place. Nor does one need to indicate the micro-processes that underlie the cognitive operations, although, in an ideal world, these would also be consistent with the postulated framework. The evidence that would argue most against the LC model would be the demonstration that part of the information was ignored.

Thus, from a psychological viewpoint, the claim that the different models capture mental processes is made at a level of analysis that represents these in an *as if* manner. Moreover, by defining the statistical properties of task environments, we have shown at a theoretical level how characteristics of models and tasks affect performance. This is an important contribution because it provides the basis for developing an environmental theory of judgmental performance (cf. Brunswik, 1952; Hammond & Stewart, 2001; Simon, 1956).

The environment, however, is not captured by statistical properties alone because context can be important. Within our framework, contextual effects are reflected in how people use decision rules. Consider, for example, what happens when cue variables are inappropriately labeled. For LC, this would reduce *ca* because less appropriate weights would be given to the variables. With the TTB model, cues might be used in an inappropriate order. In short, our approach is built on a statistical analysis of environmental tasks.

The mediating effects of context are captured by their impact on how people use decision rules.

One claim we do make is that the range of models we have considered covers the types of heuristics discussed in the literature as well, of course, as the linear model. Thus, the SV model captures what happens when people base decisions on a single cue, such as availability (Tversky & Kahneman, 1973), recognition (Goldstein & Gigerenzer, 2002), or affect (Slovic et al., 2002). All these models have in common the notion that people use a single cue that has imperfect validity. However, whether this implies that people are misguided or justified in relying on a single cue cannot be decided on an a priori basis but depends—in particular cases—on how valid the single cue is, what other relevant information is available, and the costs of making errors. It is understandable that some researchers see “the glass as half empty,” whereas others see it “as half full.”

An important contribution of our analysis is to highlight the role of error in the use of different models—as opposed to error or noise in the environment. Within LC, error is measured by the extent to which linear cognitive ability (ca or GR_s) falls short of 1.00. Here, error can have two sources: incorrect weighting of variables and inconsistency in execution. With the TTB model, the analogous error results from using variables in an inappropriate order (and, in SV, from using less valid cues). Thus, the errors in the two types of models involve both knowledge and execution, although, in the latter, execution errors are less likely given the simpler processes involved.

In future work, a more complete analysis could investigate effects of other types of processing errors on heuristic performance. For example, all our models are assumed to know the first-order correlations between cues and criterion without error. However, people typically learn this kind of information through samples of experience acquired across time. Interesting issues therefore focus on how sensitive models are to sampling variation in terms of both size and bias. One could speculate, for example, that models that need to know only the relative, as opposed to the absolute, importance of variables (e.g., SV and TTB vs. LC) would be less sensitive to sampling errors. Second, one could also model errors in the perception of cue values. Here, we suspect that models that rely on only one or two cues (e.g., SV and TTB) would be more liable to make errors.

An advantage of our meta-analysis of lens model studies is that one can say something about the effects of errors within the LC framework. Across all our studies, the mean estimates for matching (G) and consistency (R_s) are both .80 (see Table 6). Moreover, only 11% of GR_s values exceed .90. That is, the meta-analysis reveals error in both knowledge and execution, although it is an open issue as to whether these error rates are high or low. One issue they raise, however, is how much effort—say, in learning through experience and/or explicit instruction—might be needed for people to be able to outperform the better heuristics. Should people persist in using LC strategies, or should they simply seek to use the most appropriate heuristics?

We note also that although G and R_s are positively correlated, .41 ($p < .001$), neither G nor R_s is correlated with the predictability of the environment (R_e)—.06 for G and .09 for R_s . In other words, there is a trend for people to be more consistent in executing strategies when these are more valid. Perhaps more valid strategies lead to better feedback and are self-reinforcing? How-

ever, there is no relation between how predictable an environment is and people’s judgmental strategies.

An important issue we did not address in this work is the extent to which people use heuristics or the linear model in tacit (i.e., intuitive), deliberate (i.e., analytic), or even quasi-rational modes (cf. Hammond, 1996; Hogarth, 2001). The importance of this distinction is that tacit processes have little or no information-processing costs, and thus, even what may appear to be the cognitively complex operations of the LC model are not demanding. Many models of this type—or *as if* versions—were clearly used when judgmental processes have been automated such that people do not need to think about executing trade-offs. Imagine, for example, basic processes such as perception or situations in which past practice has been sufficient to hone a person’s skills. These include the judgments that most people can exercise when driving an automobile and that experts exhibit in different activities such as controlling complex systems, playing music, or even different sports (cf. Shanteau, Friel, Thomas, & Raacke, 2005). At the same time, many simple heuristics are undoubtedly tacit in nature.

An interesting feature of most tasks studied in the decision-making literature is that they are difficult precisely because people lack the experience necessary to take action without explicit thought and thus are unable to invoke valid, automatic processes. For example, the illuminating work conducted by Payne et al. (1993) demonstrated clear effort–accuracy trade-offs (involving models with different numbers of mental operations). However, these investigations were limited to relatively unfamiliar choices in which processing would have been deliberate rather than automatic. This issue emphasizes the need to understand the natural ecology of decision-making tasks (Dhimi, Hertwig, & Hoffrage, 2004). Judgmental strategies can be characterized not only by apparent analytical complexity but also by the extent to which they are executed in a tacit or deliberate manner, where the latter undoubtedly depends on the level of past experience as well as on human evolutionary heritage.

Normative Implications

Our work has many normative implications in that it spells out the conditions under which different heuristics are accurate. Moreover, the fact that this is achieved analytically—instead of through simulation—represents an advance over current practice (see also Hogarth & Karelaia, 2005a, 2006a). The analytical methods have more potential to develop results that can be generalized.

An interesting normative implication relates to the trade-offs in different types of error when using heuristics or linear models. As noted above, one way of characterizing our empirical analysis is to say that judgmental performance using the LC models is roughly equal to that of using heuristics with error, that is, of SV and TTB under random cue ordering (SV_r and TTB_r). However, is there a relation between ca and the knowledge necessary to know when and how to apply heuristic rules?

Given our results, how should a decision maker approach a predictive task? Much depends on prior knowledge of task characteristics and thus on how the individual acquired the necessary knowledge. Basically—at one extreme—if either all cues are approximately equally valid or one does not know how to weight them, EW should be used explicitly. Indeed, our results specifically demonstrate the validity of using the EW or CONF heuristics

in the absence of knowledge about the structure of the environment. Similarly—at the other extreme—when facing a noncompensatory weighting function, TTB or SV would be hard to beat with LC.

The problem lies in tasks that have more compensatory features. The key, therefore, lies in assessing *ca*. How likely is the judge to know the relative weights to give the variables? How consistent is he or she in using the judgmental strategy? On the basis of our meta-analysis, we expect that a minority of persons can meet these conditions but that much also depends on the nature of the task and the individual's experience. For example, one would be justified in trusting the judgments of the weather forecasters studied by Stewart, Roebber, and Bosart (1997) but not those of Einhorn's (1972) physicians.

Our analysis points to the importance of knowledge—about the kind of task and the capacity to handle task demands. In Table 1, we identified the levels of knowledge necessary to achieve maximum performance by all the heuristics we considered. In addition, knowledge for LC is captured by the *G* term of Equation 3. However, in many cases, people probably make judgments with less than perfect knowledge. This therefore raises important psychological issues of how people acquire such knowledge or are helped to do so. In addition, how do people encode characteristics of the environment that suggest which model to use (Rieskamp & Otto, 2006)? Overall, our results suggest that for many tasks, the errors incurred by using LC strategies are greater than those implicit in using heuristics. Thus, judgmental performance could be improved if people explicitly used appropriate heuristics instead of relying on what is often their untested and unaided judgment (see also Bröder & Schiffer, 2006). However, that people resist doing so has been documented many times (Dawes et al., 1989; Kleinmuntz, 1990). It seems that a high level of sophistication is needed to understand when to ignore information and use a heuristic. Perhaps LC strategies are psychologically attractive precisely because they allow people to feel they have considered all information (Einhorn, 1986).

Methodological Implications

Our work involves methodological innovations. Not only have we developed analytical tools for problems that frequently use simulation but also we have provided a common framework within which linear and heuristic models can be compared. This therefore opens the way to compare and contrast different ways of studying judgment and decision making.

The best predictive test of a heuristic is whether, once estimated on a sample of data, it can accurately predict a criterion in a new sample of data. The reader can therefore legitimately ask why we have not adopted this empirical strategy in our work. The reason is that there already exists evidence of successful empirical prediction by heuristics (see, e.g., Gigerenzer et al., 1999). However, these demonstrations have provided little insight as to why specific heuristics perform well and how environmental factors affect differential predictive ability. That is, they have not contributed to building appropriate theories of the environment (Brunswik, 1952; Simon, 1956). The need met by this article, therefore, is to specify how heuristics might be expected to perform under different environmental circumstances, and we believe that this issue is better framed at a theoretical level rather than relying on empirical

demonstrations alone. Given the probabilistic nature of the environment, our goal has been to create generalizable knowledge about the factors that affect heuristic performance.

It is important to point out that our theoretical approach has been tested empirically in related research (Hogarth & Karelaia, 2005a, 2006a). In that work, we used simulation to assess out-of-sample predictive accuracy and found almost perfect outcomes in repeated sampling (see also footnote 11, above). We speculated that one key to the success of heuristics is that few parameters need to be fit to the data. We noted, for example, that when regression models were also estimated from the same data, there was considerable shrinkage from fit to prediction (in excess of what one might expect from formulae for adjusted R^2). In addition to simulations, this work also made predictions for empirical data sets and found similar results, that is, in terms of both accuracy and the fact that out-of-sample predictions were better when fewer parameters needed to be estimated. We believe that an important problem for future research will be to characterize how estimating parameters of heuristics on samples of different sizes affects out-of-sample predictive ability under different environmental conditions.

Our work paid a price for analytical tractability in that there were limitations in the situations we examined. Relaxing these limitations suggests paths for further work. First, we used a binary choice paradigm involving three cues. This can be extended in two ways: to consider, first, more alternatives and, second, more cues. Our previous work (Hogarth & Karelaia, 2006a) suggests that changing the number of alternatives will not have a major influence on relative performance of different models. Increasing the number of cues, however, could have important impacts depending on the nature of intercue correlation.

Second, our analysis depends entirely on a linear model of the environment and, when looking at LC, a linear model of judgment. We believe it would be illuminating to relax these assumptions and assess the extent to which our main results change. We speculate, for example, that at a more macro level, conclusions such as the need for matching between characteristics of models of the environment and heuristics would still hold. However, given the ability of TTB to perform well across a variety of linear weighting functions, it will be instructive to see how well different heuristics perform in different, specific, nonlinear environments.

Third, all our statistical analyses have been conducted using normal distributions, and it would be of interest to see the effects of changing this assumption. In particular, what would happen in applications in which distributions are skewed and/or have fatter tails than the normal distribution? Which heuristics would have performance that is robust relative to these kinds of environmental changes and why? Further interesting complications could involve effects where models have serially correlated error terms.

Fourth, although our work has innovated in this domain by showing the effects of loss functions, we varied only the exactness parameter and not the symmetric nature of losses. It would be of interest to explore asymmetries in loss.

Fifth, as noted above, our work has identified different sources of error—in both the environment and the use of decision rules. Modeling the joint effects of such errors will be a challenging task.

Concluding Comments

This article has sought to define the environmental circumstances under which different heuristics are more or less accurate, as well as the degree of skill (linear cognitive ability) that people need to justify using linear models. An important implication of our analysis is that people do not need much computational ability to make accurate judgments but that, lacking this, they do need knowledge of when to use particular rules or heuristics. As such, the key to effective judgmental performance lies in having the knowledge necessary to guide the selection of appropriate decision rules. Important challenges for future research therefore involve both defining such knowledge explicitly and understanding how people develop this through experience.

References

- Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.
- Armeliuss, B.-A., & Armeliuss, K. (1974). The use of redundancy in multiple-cue judgments: Data from a suppressor-variable task. *American Journal of Psychology*, *87*, 385–392.
- Ashton, R. H. (1981). A descriptive study of information evaluation. *Journal of Accounting Research*, *19*, 42–61.
- Baucells, M., Carrasco, J. A., & Hogarth, R. M. (in press). Cumulative dominance and heuristic performance in multi-attribute choice. *Operations Research*.
- Brehmer, B. (1994). The psychology of linear judgement models. *Acta Psychologica*, *87*, 137–154.
- Brehmer, B., & Hagafors, R. (1986). Use of experts in complex decision making: A paradigm for the study of staff work. *Organizational Behavior and Human Decision Processes*, *38*, 181–195.
- Brehmer, B., & Joyce, C. R. B. (Eds.). (1988). *Human judgment: The SJT view*. Amsterdam: North-Holland.
- Brehmer, B., & Kuylensstierna, J. (1980). Content and consistency in probabilistic inference tasks. *Organizational Behavior and Human Performance*, *26*, 54–64.
- Bröder, A. (2000). Assessing the empirical validity of the “take-the-best” heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1332–1346.
- Bröder, A. (2003). Decision making with the “adaptive toolbox”: Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 611–625.
- Bröder, A., & Schiffer, S. (2003). Take the best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, *132*, 277–293.
- Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 904–918.
- Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: University of Chicago Press.
- Camerer, C. F. (1981). General conditions for the success of bootstrapping models. *Organizational Behavior and Human Performance*, *27*, 411–422.
- Chasseigne, G., Grau, S., Mullet, E., & Cama, V. (1999). How well do elderly people cope with uncertainty in a learning task? *Acta Psychologica*, *103*, 229–238.
- Chasseigne, G., Mullet, E., & Stewart, T. R. (1997). Aging and multiple cue probability learning: The case of inverse relationships. *Acta Psychologica*, *97*, 235–252.
- Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. New York: Academic Press.
- 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, March 31). Clinical versus actuarial judgment. *Science*, *243*, 1668–1674.
- Deane, D. H., Hammond, K. R., & Summers, D. A. (1972). Acquisition and application of knowledge in complex inference tasks. *Journal of Experimental Psychology*, *92*, 20–26.
- Dhimi, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*, 959–988.
- Doherty, M. E., Tweney, R. D., O'Connor, R. M., Jr., & Walker, B. (1988). *The role of data and feedback error in inference and prediction: Final report for ARI contract MDA903-85-K-0193*. Bowling Green, OH: Bowling Green State University.
- Einhorn, H. J. (1972). Expert measurement and mechanical combination. *Organizational Behavior and Human Performance*, *7*, 86–106.
- Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment*, *50*, 387–395.
- Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, *13*, 171–192.
- Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process tracing models of judgment. *Psychological Review*, *86*, 465–485.
- Fasolo, B., McClelland, G. H., & Todd, P. M. (2007). Escaping the tyranny of choice: When fewer attributes make choice easier. *Marketing Theory*, *7*, 13–26.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, *103*, 592–596.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Goldberg, L. R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical judgment. *Psychological Bulletin*, *73*, 422–432.
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75–90.
- Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, *62*, 255–262.
- Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York: Oxford University Press.
- Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review*, *71*, 438–456.
- Hammond, K. R., & Stewart, T. R. (Eds.). (2001). *The essential Brunswik: Beginnings, explications, applications*. New York: Oxford University Press.
- Hammond, K. R., & Summers, D. A. (1965). Cognitive dependence on linear and nonlinear cues. *Psychological Review*, *72*, 215–224.
- Hammond, K. R., Summers, D. A., & Deane, D. H. (1973). Negative effects of outcome-feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance*, *9*, 30–34.
- Hammond, K. R., Wilkins, M. M., & Todd, F. J. (1966). A research paradigm for the study of interpersonal learning. *Psychological Bulletin*, *65*, 221–232.
- Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, *112*, 494–508.
- Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, *57*, 116–131.
- Hoffman, P. J., Earle, T. C., & Slovic, P. (1981). Multidimensional functional learning (MFL) and some new conceptions of feedback. *Organizational Behavior and Human Performance*, *27*, 75–102.

- Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, *90*, 197–217.
- Hogarth, R. M. (1987). *Judgement and choice: The psychology of decision* (2nd ed.). Chichester, England: Wiley.
- Hogarth, R. M. (2001). *Educating intuition*. Chicago: University of Chicago Press.
- Hogarth, R. M., Gibbs, B. J., McKenzie, C. R. M., & Marquis, M. A. (1991). Learning from feedback: Exactingness and incentives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 734–752.
- Hogarth, R. M., & Karelaia, N. (2005a). Ignoring information in binary choice with continuous variables: When is less “more”? *Journal of Mathematical Psychology*, *49*, 115–124.
- Hogarth, R. M., & Karelaia, N. (2005b). Simple models for multi-attribute choice with many alternatives: When it does and does not pay to face trade-offs with binary attributes. *Management Science*, *5*, 1860–1872.
- Hogarth, R. M., & Karelaia, N. (2006a). Regions of rationality: Maps for bounded agents. *Decision Analysis*, *3*, 124–144.
- Hogarth, R. M., & Karelaia, N. (2006b). “Take-the-best” and other simple strategies: Why and when they work “well” in binary choice. *Theory and Decision*, *61*, 205–249.
- Holzworth, R. J., & Doherty, M. E. (1976). Feedback effects in a metric multiple-cue probability learning task. *Bulletin of the Psychonomic Society*, *8*, 1–3.
- Hursch, C. J., Hammond, K. R., & Hursch, J. L. (1964). Some methodological considerations in multiple-probability studies. *Psychological Review*, *71*, 42–60.
- Jarnecke, R. W., & Rudestam, K. E. (1976). Effects of amounts and units of information on the judgmental process. *Perceptual and Motor Skills*, *13*, 823–829.
- Juslin, P., & Persson, M. (2002). PROBABILITIES from EXemplars (PROBEX): A “lazy” algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, *26*, 563–607.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582–591.
- Karelaia, N. (2006). Thirst for confirmation in multi-attribute choice: Does search for consistency impair decision performance? *Organizational Behavior and Human Decision Processes*, *100*, 128–143.
- Karelaia, N., & Hogarth, R. M. (2007). *Determinants of linear judgment: A meta-analysis of lens model studies* (DEE Working Paper No. 1007). Barcelona, Spain: Universitat Pompeu Fabra.
- Katsikopoulos, K. V., & Martignon, L. (2006). Naïve heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology*, *50*, 488–494.
- Keeney, R., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value trade-offs*. New York: Wiley.
- Kessler, L., & Ashton, R. H. (1981). Feedback and prediction achievement in financial analysis. *Journal of Accounting Research*, *19*, 146–162.
- Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211–228.
- Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, *107*, 296–310.
- Lafon, P., Chasseigne, G., & Mullet, E. (2004). Functional learning among children, adolescents, and young adults. *Journal of Experimental Child Psychology*, *88*, 334–347.
- Lee, J.-W., & Yates, J. F. (1992). How quantity judgment changes as the number of cues increases: An analytical framework and review. *Psychological Bulletin*, *112*, 363–377.
- Lindell, M. K. (1976). Cognitive and outcome feedback in multiple-cue probability learning tasks. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 739–745.
- Louvière, J. J. (1988). *Analyzing decision making: Metric conjoint analysis*. Thousand Oaks, CA: Sage.
- Luce, M. F., Payne, J. W., & Bettman, J. R. (1999). Emotional trade-off difficulty and choice. *Journal of Marketing Research*, *36*, 143–159.
- Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd, & 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.
- McKenzie, C. R. M. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology*, *26*, 209–239.
- Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.
- Montgomery, H. (1983). Decision rules and the search for dominance structure: Towards a process model of decision making. In P. Humphreys, O. Svenson, & A. Vari (Eds.), *Analyzing and aiding decision processes* (pp. 343–369). Amsterdam: North-Holland.
- Muchinsky, P. M., & Dudycha, A. L. (1975). Human inference behavior in abstract and meaningful environments. *Organizational Behavior and Human Performance*, *13*, 377–391.
- Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing “one-reason” decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 53–65.
- Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone “takes-the-best.” *Organizational Behavior and Human Decision Processes*, *91*, 82–96.
- O’Connor, M., Remus, W., & Lim, K. (2005). Improving judgmental forecasts with judgmental bootstrapping and task feedback support. *Journal of Behavioral Decision Making*, *18*, 246–260.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.
- Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 141–167). New York: Oxford University Press.
- Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*, 207–236.
- Rothstein, H. G. (1986). The effects of time pressure on judgment in multiple cue probability learning. *Organizational Behavior and Human Decision Processes*, *37*, 83–92.
- Shanteau, J., Friel, B. M., Thomas, R. P., & Raacke, L. (2005). Development of expertise in a dynamic decision-making environment. In T. Betsch & S. Haberstroh (Eds.), *The routines of decision making* (pp. 251–270). Mahwah, NJ: Erlbaum.
- Shanteau, J., & Thomas, R. P. (2000). Fast and frugal heuristics: What about unfriendly environments? *Behavioral and Brain Sciences*, *23*, 762–763.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99–118.
- Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review*, *63*, 129–138.
- Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). New York: Cambridge University Press.
- Steinmann, D. O. (1974). Transfer of lens model learning. *Organizational Behavior and Human Performance*, *12*, 1–16.
- Steinmann, D. O., & Doherty, M. E. (1972). A lens model analysis of a bookbag and poker chip experiment: A methodological note. *Organizational Behavior and Human Performance*, *8*, 450–455.

- Stewart, T. R., Roebber, P. J., & Bosart, L. F. (1997). The importance of the task in analyzing expert judgment. *Organizational Behavior and Human Decision Processes*, 69, 205–219.
- Summers, S. A., Summers, R. C., & Karkau, V. T. (1969). Judgments based on different functional relationships between interacting cues and a criterion. *American Journal of Psychology*, 82, 203–211.
- Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, 25, 219–225.
- Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hirsch, Hammond, and Hirsch and by Hammond, Hirsch, and Todd. *Psychological Review*, 71, 528–530.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 4, 207–232.
- Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293–315.
- Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, 83, 213–217.
- York, K. M., Doherty, M. E., & Kamouri, J. (1987). The influence of cue unreliability in a multiple cue probability learning task. *Organizational Behavior and Human Decision Processes*, 39, 303–317.
- Youmans, R. J., & Stone, E. R. (2005). To thy own self be true: Finding the utility of cognitive information feedback. *Journal of Behavioral Decision Making*, 18, 319–341.

Appendix A

The Expected Accuracies of LC, EW, CONF, and TTB

The LC Model

Following the same rationale as the single variable (SV) model, we can also determine the probability that using a linear combination (LC) of cues will result in a correct choice. That is, expressing Y_{ea} and Y_{eb} as functions of Y_{sa} and Y_{sb} , define appropriate error terms, ω_a and ω_b , and substitute $\rho_{Y_e Y_s}$ for $\rho_{Y_e X_1}$, and Y_{sa} and Y_{sb} for X_{1a} and X_{1b} , respectively. Thus, $2P\{(Y_{ea} > Y_{eb}) \cap (Y_{sa} > Y_{sb})\}$ can also be found through Equation 7, with $f(\underline{d})$ defined as in SV. The only difference between SV and LC lies in the variance–covariance matrix, M_f , that, for the LC model, is

$$M_{f_{LC}} = \begin{pmatrix} 2 & 2\rho_{Y_e Y_s} \\ 2\rho_{Y_e Y_s} & 2 \end{pmatrix}.$$

The EW Model

Equal weighting (EW) is, of course, a special case of LC. Define $d_2 = \bar{X}_a - \bar{X}_b$, where $\bar{X}_a = \frac{1}{k} \sum_{j=1}^k X_{ja}$ and $\bar{X}_b = \frac{1}{k} \sum_{j=1}^k X_{jb}$, and note that d_2 is a normal variable with a mean of 0. (The variable d_1 for EW is the same as for LC: $d_1 = Y_{ea} - Y_{eb}$.) Thus, the expected accuracy of EW can be defined by Equation 7, taking into consideration that the appropriate variance–covariance matrix is

$$M_{f_{EW}} = \begin{pmatrix} 2 & 2\rho_{Y_e \bar{X}} \sigma_{\bar{X}} \\ 2\rho_{Y_e \bar{X}} \sigma_{\bar{X}} & 2\sigma_{\bar{X}}^2 \end{pmatrix}.$$

(Note that from Equation 3, it follows that $\rho_{Y_e \bar{X}} = \rho_{Y_e \bar{X}} R_e$, assuming $\rho_{\epsilon_e \epsilon_s} = 0$.)

The CONF Model

CONF examines cues sequentially and makes a choice when two cues favoring one alternative are encountered. Therefore, this model selects the better alternative out of two with the probability given by

$$2 \left[\begin{array}{l} P\{(Y_{ea} > Y_{eb}) \cap (X_{1a} > X_{1b}) \cap (X_{2a} > X_{2b})\} + \\ P\{(Y_{ea} > Y_{eb}) \cap (X_{1a} > X_{1b}) \cap (X_{2a} < X_{2b}) \cap (X_{3a} > X_{3b})\} + \\ P\{(Y_{ea} < Y_{eb}) \cap (X_{1a} > X_{1b}) \cap (X_{2a} < X_{2b}) \cap (X_{3a} < X_{3b})\} \end{array} \right] \\ = 2 \left[\int_0^\infty \int_0^\infty \int_0^\infty f_1(\underline{d}) \underline{d} \underline{d} \underline{d} + \int_0^\infty \int_0^\infty \int_{-\infty}^0 \int_0^\infty f_2(\underline{d}) \underline{d} \underline{d} \underline{d} + \int_{-\infty}^0 \int_0^\infty \int_{-\infty}^0 \int_0^\infty f_2(\underline{d}) \underline{d} \underline{d} \underline{d} \right], \quad (A1)$$

where both $f_1(\underline{d}) = f_1(d_1, d_2, d_3)$ and $f_2(\underline{d}) = f_2(d_1, d_2, d_3, d_4)$ are the normal multivariate probability density functions, the variance–covariance matrix specific to each being

$$M_{f_1} = \begin{pmatrix} 2 & 2\rho_{Y_e X_1} & 2\rho_{Y_e X_2} \\ 2\rho_{Y_e X_1} & 2 & 2\rho_{X_1 X_2} \\ 2\rho_{Y_e X_2} & 2\rho_{X_1 X_2} & 2 \end{pmatrix}$$

(Appendixes continue)

and

$$M_{f_2} = \begin{pmatrix} 2 & 2\rho_{Y_e X_1} & 2\rho_{Y_e X_2} & 2\rho_{Y_e X_3} \\ 2\rho_{Y_e X_1} & 2 & 2\rho_{X_1 X_2} & 2\rho_{X_1 X_3} \\ 2\rho_{Y_e X_2} & 2\rho_{X_1 X_2} & 2 & 2\rho_{X_2 X_3} \\ 2\rho_{Y_e X_3} & 2\rho_{X_1 X_3} & 2\rho_{X_2 X_3} & 2 \end{pmatrix}.$$

The TTB Model

The take-the-best (TTB) model also assesses cues sequentially. It makes a choice when a discriminating cue is found. In this article, we consider TTB with a fixed threshold $t (>0)$. Thus, the model stops consulting cues and makes a decision when $|x_{ia} - x_{ib}| > t$. This involves both cases when $x_{ia} - x_{ib} > t$ and cases when $x_{ib} - x_{ia} > t$. Because the two cases are symmetric, the probability that TTB selects the better alternative is

$$\begin{aligned} & 2 \left[\begin{aligned} & P\{(Y_{ea} > Y_{eb}) \cap (X_{1a} - X_{1b} \geq t)\} + \\ & P\{(Y_{ea} > Y_{eb}) \cap (|X_{1a} - X_{1b}| < t) \cap (X_{2a} - X_{2b} \geq t)\} + \\ & P\{(Y_{ea} > Y_{eb}) \cap (|X_{1a} - X_{1b}| < t) \cap (|X_{2a} - X_{2b}| < t) \cap (X_{3a} - X_{3b} \geq t)\} \end{aligned} \right] \\ & + P\{(Y_{ea} > Y_{eb}) \cap (|X_{1a} - X_{1b}| < t) \cap (|X_{2a} - X_{2b}| < t) \cap (|X_{3a} - X_{3b}| < t)\} \\ & = 2 \left[\int_0^\infty \int_t^\infty f_3(\underline{d}) d\underline{d} + \int_0^\infty \int_{-t}^t \int_t^\infty f_1(\underline{d}) d\underline{d} + \int_0^\infty \int_{-t}^t \int_{-t}^t \int_t^\infty f_2(\underline{d}) d\underline{d} \right] + \int_0^\infty \int_{-t}^t \int_{-t}^t \int_{-t}^t f_3(\underline{d}) d\underline{d}, \quad (A2) \end{aligned}$$

where both $f_1(\underline{d}) = f_1(d_1, d_2, d_3)$ and $f_2(\underline{d}) = f_2(d_1, d_2, d_3, d_4)$ are the same as in CONF and $f_3(\underline{d}) = f_3(d_1, d_2)$ is defined similarly, using the appropriate variance-covariance matrix:

$$M_{f_3} = \begin{pmatrix} 2 & 2\rho_{Y_e X_1} \\ 2\rho_{Y_e X_1} & 2 \end{pmatrix}.$$

Appendix B

The Expected Loss of the CONF and Take-the-Best (TTB) Models

The expected loss of CONF is

$$\begin{aligned} & 2L \left[\begin{aligned} & P\{(Y_{ea} < Y_{eb}) \cap (X_{1a} > X_{1b}) \cap (X_{2a} > X_{2b})\} + \\ & P\{(Y_{ea} < Y_{eb}) \cap (X_{1a} > X_{1b}) \cap (X_{2a} < X_{2b}) \cap (X_{3a} > X_{3b})\} + \\ & P\{(Y_{ea} > Y_{eb}) \cap (X_{1a} > X_{1b}) \cap (X_{2a} < X_{2b}) \cap (X_{3a} < X_{3b})\} \end{aligned} \right] \\ & = 2\alpha \left[\int_{-\infty}^0 \int_0^\infty \int_0^\infty d^2 f_1(\underline{d}) d\underline{d} + \int_{-\infty}^0 \int_0^\infty \int_0^\infty d^2 f_2(\underline{d}) d\underline{d} + \int_0^\infty \int_0^\infty \int_{-\infty}^0 \int_{-\infty}^0 d^2 f_2(\underline{d}) d\underline{d} \right], \quad (B1) \end{aligned}$$

where $f_1(\underline{d})$ and $f_2(\underline{d})$ are as defined in Appendix A.

The expected loss of TTB is

$$\begin{aligned} & 2L \left[\begin{aligned} & P\{(Y_{ea} < Y_{eb}) \cap (X_{1a} - X_{1b} \geq t)\} + \\ & P\{(Y_{ea} < Y_{eb}) \cap (|X_{1a} - X_{1b}| < t) \cap (X_{2a} - X_{2b} \geq t)\} + \\ & P\{(Y_{ea} < Y_{eb}) \cap (|X_{1a} - X_{1b}| < t) \cap (|X_{2a} - X_{2b}| < t) \cap (X_{3a} - X_{3b} \geq t)\} \end{aligned} \right] \\ & + LP\{(Y_{ea} < Y_{eb}) \cap (|X_{1a} - X_{1b}| < t) \cap (|X_{2a} - X_{2b}| < t) \cap (|X_{3a} - X_{3b}| < t)\} \\ & = \alpha \left(2 \left[\int_{-\infty}^0 \int_t^\infty \int_t^\infty d^2 f_3(\underline{d}) d\underline{d} + \int_{-\infty}^0 \int_{-t}^t \int_t^\infty d^2 f_1(\underline{d}) d\underline{d} + \int_{-\infty}^0 \int_{-t}^t \int_{-t}^t \int_t^\infty d^2 f_2(\underline{d}) d\underline{d} \right] \right. \\ & \left. + \int_{-\infty}^0 \int_{-t}^t \int_{-t}^t \int_{-t}^t d^2 f_2(\underline{d}) d\underline{d} \right), \quad (B2) \end{aligned}$$

where $f_1(d)$, $f_2(d)$, and $f_3(d)$ are as defined in Appendix A.

Appendix C
Selected Three-Cue Studies

Environment & study	Task	Number of conditions/tasks	Total number of participants	Stimuli per participant	R_e across conditions (range)	Mean human performance across conditions (range)	
						r_a	GR_s
Equal weighting environments							
1. Ashton (1981)	Predicting prices	3	138	30	0.01–0.98	–0.17–0.19	0.01–0.87
2a. Brehmer & Hagafors (1986)	Artificial prediction task	1	10	15	1.00	0.97	0.95
3. Chasseigne, Grau, Mullet, & Cama (1999)	Artificial prediction task	5	220	120	0.57–0.98	0.37–0.78	0.67–0.82
Compensatory environments							
4. Holzworth & Doherty (1976)	Artificial prediction task	6	58	25	0.71	0.54–0.64	0.76–0.91
5. Chasseigne, Mullet, & Stewart (1997)–Experiment 1	Artificial prediction task	6	96	26	0.96	0.34–0.70	0.35–0.73
6. Kessler & Ashton (1981)	Prediction of corporate bond ratings	4	69	34	0.74	0.52–0.64	0.71–0.88
7a. Steinmann (1974)	Artificial prediction task	9	11	300	0.63–0.78	0.45–0.57	0.68–0.84
Noncompensatory environments							
2b. Brehmer & Hagafors (1986)	Artificial prediction task	2	20	15	0.77–1.00	0.74–0.78	0.71–0.75
8. Deane, Hammond, & Summers (1972)–Experiment 2	Artificial prediction task	2	40	20	0.94	0.59–0.84	0.65–0.89
9. Hammond, Summers, & Deane (1973)	Artificial prediction task	3	30	20	0.92	0.05–0.78	0.14–0.83
10. Hoffman, Earle, & Slovic (1981)	Artificial prediction task	9	182	25	0.94	0.09–0.71	0.15–0.78
7b. Steinmann (1974)	Artificial prediction task	6	11	100	0.63–0.74	0.44–0.65	0.70–0.85
11. O'Connor, Remus, & Lim (2005)	Artificial prediction task	4	77	20	0.81–0.84	0.59–0.72	0.71–0.87
12. Youmans & Stone (2005)	Prediction of income levels	4	117	50	0.44	0.35–0.42	0.88–0.97
Total		64	1,079				

Note. All studies reported involved between-subject designs except for Studies 7a and 7b. Three studies—8, 9, and 10—were said to have identical parameters. However, there must have been some rounding differences because of marginally different values reported for R_e .

(Appendixes continue)

Appendix D
Selected Two-Cue Studies

Environment & study	Task	Number of conditions/tasks	Total number of participants	Stimuli per participant	R_e across conditions (range)	Mean human performance across conditions (range)	
						r_a	GR_s
Equal weighting environments							
1. Jarnecke & Rudestam (1976)	Predict academic achievement	1	15	50	0.42	0.28	0.71
2. Lafon, Chasseigne, & Mullet (2004)	Artificial prediction task	4	439	30	0.96	0.00–0.90	0.00–0.94
3. Rothstein (1986)	Artificial prediction task	6	72	100	1.00	0.81–1.00	0.80–1.00
4. Summers, Summers, & Karkau (1969)	Judging the age of blood cells	1	16	64	0.99	0.73	0.73
5. Brehmer & Kuylenstierna (1980)	Artificial prediction task	5	40	15	0.57–0.81	0.38–0.76	0.67–0.90
Noncompensatory environments							
6. Armelius & Armelius (1974)	Artificial prediction task	3	63	25	0.99–1.00	0.32–0.96	0.32–0.95
7. Doherty, Tweney, O'Connor, & Walker (1988)							
Experiment 2	Artificial prediction task	3	45	25	0.79–1.00	0.70–0.73	0.74–0.92
Experiment 6	Artificial prediction task	2	30	50	0.87–1.00	0.53–0.66	0.58–0.73
8. Hammond & Summers (1965)	Artificial prediction task	3	30	20	0.71	0.49–0.85	0.48–0.59
9. Lee & Yates (1992)	Postdicting student success	2	40	NA	0.38	0.24–0.29	0.51–0.59
10. Muchinsky & Dudycha (1975)							
Experiment 1	Artificial prediction task	2	160	150	0.72	0.04–0.30	0.11–0.54
Experiment 2	Artificial prediction task	2	160	150	0.96	0.03–0.45	0.01–0.32
11. Steinmann & Doherty (1972)	Assessing subjective probabilities in a bookbag and poker chip task	1	22	192	0.95	0.67	0.70
12. York, Doherty, & Kamouri (1987)	Artificial prediction task	3	45	25	0.86	0.53–0.64	0.62–0.74
Total		38	1,177				

Note. The numbers of participants in Studies 3 and 9 are approximations because this information is not available. In Study 11, human performance was measured through medians.

Received November 29, 2006

Revision received March 12, 2007

Accepted March 13, 2007 ■