On heuristic and linear models of judgment: Mapping the demand for knowledge*

Robin M. Hogarth[1] & Natalia Karelaia[2]
ICREA & Universitat Pompeu Fabra[1], Barcelona,
HEC Université de Lausanne[2], Lausanne

robin.hogarth@upf.edu
natalia.karelaia@unil.ch

June 18, 2006

* This research was financed partially by grants from the Spanish Ministerio de Educación y Ciencia (Hogarth) and the Swiss National Science Foundation (Karelaia). The authors have benefited from comments received at the Brunswik Society 2005 Annual Meeting in Toronto, and seminars at the University of Basel, INSEAD, and the Max Planck Institute, Berlin. We are particularly indebted to Thomas Stewart, Michael Doherty, and the library at Universitat Pompeu Fabra for helping us locate many lens model studies. For correspondence, please contact Robin M. Hogarth at Universitat Pompeu Fabra, Department of Economics and Business, Ramon Trias Fargas 25-27, 08005, Barcelona, Spain. (Tel: +34 93 542 2561, Fax + 34 93 542 1746).

Abstract

Research on judgment and decision making presents a confusing picture of human abilities. For example, much research has emphasized the dysfunctional aspects of judgmental heuristics, and yet, other findings suggest that these can be highly effective. A further line of research has modeled judgment as resulting from "as if" linear models. This paper illuminates the distinctions in these approaches by providing a common analytical framework based on the central theoretical premise that understanding human performance requires specifying how characteristics of the decision rules people use interact with the demands of the tasks they face. Our work synthesizes the analytical tools of "lens model" research with novel methodology developed to specify the effectiveness of heuristics in different environments and allows direct comparisons between the different approaches. We illustrate with both theoretical analyses and simulations. We further link our results to the empirical literature by a meta-analysis of lens model studies and estimate both human and heuristic performance in the same tasks. Our results highlight the trade-off between linear models and heuristics. Whereas the former are cognitively demanding, the latter are simple to use. However, they require knowledge – and thus "maps" – of when and which heuristic to employ.

Keywords: Decision making; heuristics; linear models; lens model; judgmental biases.

JEL classification: D81, M10.

Two classes of models have dominated research on judgment and decision making over the last decades. In one, explicit recognition is given to the costs and limits of information processing and people are assumed to use simplifying heuristics – typically making use of only part of the information available (Kahneman, Slovic, & Tversky, 1982; Gigerenzer, Todd, & the ABC Research Group, 1999). 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).

Research on these models has been conducted within different traditions with few attempts to unify the two approaches (however, see Hammond, 1990). Whereas such unification is not our goal, we recognize the validity of both approaches and seek to illuminate their complementarities. For example, recent research suggests that people can process information in distinctive ways (cf., Chaiken & Trope, 1999), variously described as "experiential" vs. "rational" (Epstein, 1994), "System 1" vs. "System 2" (Stanovich & West, 1998), or "tacit" vs. "deliberate" (Hogarth, 2001). The former denote processes that are intuitive or heuristic whereas the latter are the outcomes of more deliberative processes. We do not propose a one-to-one correspondence between the dual process approach, on the one hand, and heuristic and algebraic models, on the other hand. However, the analogy emphasizes the advantages of seeking complementarities.

The topic of heuristics has been central to research on judgment and decision making and 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) in everyday life, many questions are still unresolved. One important set of issues centers on understanding the relative efficacy of different heuristics and, in particular, explicating the environmental conditions when these are effective.

At one level, this failure is surprising in that Herbert Simon – whose work is held in high esteem by researchers with differing views about heuristics – specifically emphasized the importance of environmental factors. In particular, 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).

Interest, however, of most research on heuristics has centered on specific rules such as representativeness (Kahneman & Tversky, 1972), availability (Tversky & Kahneman, 1973), recognition (Goldstein & Gigerenzer, 2002), and affect (Slovic, Finucane, Peters, & MacGregor, 2002) that limit information processing costs and there have been few attempts to understand possible environmental effects.[1]

At the same time that Simon was publishing his seminal work on heuristics, the use of algebraic, and particularly 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).[2] Contrary to work on heuristics, this research has shown concern for environmental factors. Specifically, by depicting Brunswik's (1952) lens model within a linear framework, Hammond and his colleagues were able to describe psychological

[1] As stated ironically by one of our colleagues, it is as though researchers on heuristics suffer collectively from the "fundamental attribution error" (Ross, 1977) whereby explanations of behavior fail to take environmental factors into consideration.

[2] The earliest representation of judgment as a linear model that we know of goes back to Wallace (1923).

achievement in the form of an equation – the lens model equation – that captures effects of both individuals and the environment (Hammond, Hursch, & Todd, 1964; Hursch, Hammond, & Hursch, 1964; Tucker, 1964). 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 "work-horse" of judgment and decision making research from both descriptive and prescriptive viewpoints. As to the latter, consider the influence of linear models in multi-attribute theory (see, e.g., Keeney & Raiffa, 1976) as well as the literatures on bootstrapping (Goldberg, 1970; Camerer, 1981; Russo & Schoemaker, 2002), equal-weighting (Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Wainer, 1976; Dawes, 1979), and the statistical-clinical debate (Meehl, 1954; Dawes, Faust, & Meehl, 1989; Kleinmuntz, 1990).

However, despite the ubiquity of the linear model in representing human information integration, its psychological validity has been questioned. First, when the amount of information exceeds a threshold (e.g., 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 (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 effective are these relative to the more cognitively demanding linear model? Moreover, if heuristics neglect information and/or avoid trade-offs, how do these features contribute to their success or failure, and when?

Our purpose is to illuminate these and related issues 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. Thus, if the decision maker weights the cues appropriately (using a linear model) she will achieve the maximum predictive performance. However (as we explain below), this is an exacting standard to achieve. Thus, what are the consequences of abandoning the linear rule and using simpler heuristics? Moreover, when will different heuristics perform relatively well or badly?

Specifically, we consider five models and, to simplify the analysis, only consider three cues. (We return to this issue in the Discussion.) Two of these models are linear and three are heuristics. Whereas we could have chosen many variations of these models, we believe they are sufficient to illustrate our approach.

First, we consider what happens when the decision maker can be modeled as if she were using a linear combination of the cues (LC) with respect to the weights applied to the variables and is also inconsistent (cf., Hoffman, 1960). Note carefully that we are not saying that the decision maker actually uses a linear formula but can be modeled "as if." We justify this approach on the grounds that linear models can often provide higher-level representations of underlying processes such that their outcomes are consistent with a variety of different models (for further elaboration, see Einhorn et al., 1979). Moreover, when the amount of information to be integrated is limited, the linear model can also provide a good process description (Payne, Bettman, & Johnson, 1993).

Second, the decision maker is unable to differentiate the weights that should be given to the variables and simplifies by giving equal weight to each (EW).[3] EW, of course, is a special case of LC and has been demonstrated to have desirable properties (Dawes & Corrigan, 1974).

Third, the decision maker uses the "take-the-best" (TTB) heuristic proposed by Gigerenzer and Goldstein (1996). This works as follows. It is first assumed that the decision maker can order attributes or cues 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, Todd, & the ABC Research Group, 1999).[4]

There is experimental evidence that people use TTB-like strategies, although not exclusively (Rieskamp & Hoffrage, 1999; 2002; Bröder, 2000; 2003; Bröder & Schiffer, 2003; Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003). 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, 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

---

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

[4] In Gigerenzer and Goldstein's (1996) formulation, TTB operates on cues that can only 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, i.e., a cue only discriminates between two alternatives if the difference between the values of the cues exceeds a specified value t (>0).

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. With three cues, then, CONF only requires that two cues favor the chosen alternative. Moreover, CONF has the advantage that choice is insensitive to the order in which cues are consulted. Thus, 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. This therefore also models any heuristic that is based on a single variable such as in judgments by representativeness (Kahneman & Tversky, 1972), availability (Tversky & Kahneman, 1973), recognition (Gigerenzer & Goldstein, 2002), or affect (Slovic et al., 2002). In these latter cases, however, the variable would not necessarily be observable by a third party but would represent an intuitive feeling or judgment experienced by the decision maker in the situation (e.g., an assessment of similarity, knowledge of recognition, or a feeling of liking).

Insert Figure 1 about here

It is important to note that all these rules represent feasible psychological processes. Figure 1 specifies and compares what needs to be known for each of the models to achieve its *maximum* performance. As can be seen, this can be decomposed between knowledge about the specific cue values taken by the three variables under consideration (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

[5] 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.

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 needs to identify the cue with the greatest validity (if there is more than one). Neither EW nor CONF requires knowledge about weights.

Whereas it is difficult to tell whether obtaining values of cue variables or knowing something about how cues vary in importance is more taxing cognitively, we have attempted an ordering of the models in Figure 1 from most to least taxing. Clearly, LC is the most taxing and, as noted above, the important issue to understand is how sensitive it is 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, Hursch, & Todd, 1964; Hursch, Hammond, & Hursch, 1964; Tucker, 1964; Hammond & Stewart, 2001) combined with more recent analytic methods that were developed to determine the performance of heuristic decision rules (Hogarth & Karelaia, 2005a; in press; 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 (cf. Figure 1), we see our work as mapping the demand for knowledge in different regions of the environment. In other words, to make effective 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 the type of weighting function (i.e., how the environment weights cues); cue inter-correlation; the predictability of the environment; and loss functions. Whereas the weighting function determines which heuristic is best suited to specific tasks, the other factors moderate the advantages of selecting the correct rule. Both cue redundancy (i.e., inter-correlation) and noise (i.e., lack of predictability) reduce differences between model performance but these can be augmented or diminished according to the loss function used. We also show that "sensible" models make identical predictions in more cases than might have been imagined a priori. However, since they disagree across 8-30% of the cases we examined, 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 or better than heuristics in binary choice. We find that performance using LC rules generally falls short of that of appropriate heuristics unless decision makers have high "linear cognitive ability" (which we quantify). This analysis is supported by a meta-analysis of lens model studies in which we estimate linear cognitive ability across some 250 tasks and also demonstrate that, within the same tasks, individuals vary in their ability to outperform heuristics using LC models.

This paper is organized as follows. We first briefly review literature that has considered the effectiveness of heuristic decision models. For the most part, this has been dependent on empirical demonstrations and simulations and, as such, conclusions cannot be easily generalized. In contrast, our approach, developed in the subsequent section, is based on statistical theory. This allows us to make theoretical predictions of model accuracy in terms of both percentage correct predictions and expected losses. To facilitate the exposition, we present the underlying rationale with respect to the SV, LC, and EW models in the main text and the equations for the other models in Appendices A and B. 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 meta-analysis of the lens model literature and leads to the conclusions summarized above. Finally, we consider psychological, normative, and methodological implications of our work as well as suggestions for future research.

Evidence on the effectiveness of simple, heuristic models

Interest in the use of heuristic decision models has fueled much research (and controversy) in judgment and decision making. The initial impetus from Simon's work on bounded rationality (Simon, 1955; 1956) was to emphasize the need for humans to use heuristic methods (or to "satisfice") because of inherent cognitive limitations. Moreover, Simon stressed the importance of understanding how the structure of the environment affects the relative effectiveness 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, Slovic, & Tversky, 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 were or were not effective (cf., Hogarth, 1981).

Nonetheless, the positive side of heuristic use has also been emphasized. (Although, here too a concern for explicating environmental limitations has not been paramount.) One line of research has emphasized equal-weighting models, the effectiveness of which was demonstrated through simulations and empirical examples (Dawes & Corrigan, 1974; Dawes, 1979). In further simulations, Payne, Bettman and Johnson (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 models 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.)

The predictive effectiveness of TTB was first demonstrated by Gigerenzer and Goldstein (1996) in an empirical illustration and then subsequently replicated over 18 further datasets (Gigerenzer, Todd, et al., 1999). Specifically, these studies showed that TTB predicted more accurately (on cross-validation) than EW and multiple regression when the criterion was 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 (in press) in a simulation of multi-attribute 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 inter-correlations among the attributes and types of weighting functions. Results showed important effects for both. With differential weighting, one attribute was sufficient to capture at least 90% of total utility. With positive inter-correlation among attributes, there was little difference between equal and differential weighting. With negative inter-correlation, however, equal weighting was sensitive to the number of attributes used (the more, the better).

Despite these 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 effective. Some work has, however, considered specific cases. Einhorn and Hogarth (1975), for example, provided a theoretical rationale for the effectiveness of equal weighting relative to multiple regression. Martignon and Hoffrage (1999; 2002) and Katsikopoulos and Martignon (in press) explored the conditions under which TTB or equal weighting should be preferred in binary choice. Hogarth and Karelaia (2005a; in press, a) and Baucells, Carrasco, and Hogarth (2006) have examined why TTB and other simple models perform well with binary attributes in error-free environments.

Finally, in related work (Hogarth & Karelaia, 2005b; in press, b), we have provided an analytical framework for determining what we named "regions of rationality," i.e., the specification of when heuristic models are and are not effective. The current paper builds on these foundations.

To facilitate presentation of our analytical results, we first briefly explain the logic of the lens model and the so-called "lens model equation" (Tucker, 1964). We then derive equations for the predictive ability of the heuristics we examine in terms of expected predicted correct in binary choice as well as squared-error loss functions. Our strategy involves presenting the key ideas in the main text with details provided in appendices. An important difference between studies of heuristic judgment and those using the LC framework (or lens model) is that the empirical criterion for the latter – known as “achievement” – is framed within the context of the correlation between judgments and outcomes as opposed to percentage correct predictions in binary choice. In comparing paradigms, therefore, we transform correlational achievement into equivalent percentage correct in binary choice.

Theoretical development

To motivate the theoretical development, imagine a binary choice situation 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 variable, i.e., a measure of subsequent job performance, can be observed at a later date and that a correct decision was taken if the criterion is greater for the chosen candidate.[6] Denote the criterion by the random variable Y_e such that if A happened to be the correct choice, one would observe $y_{ea} > y_{eb}$.[7]

Within the lens model framework – see Figure 2 – 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). These equations are, respectively:

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

and

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

---

[6] In practice one would typically only be able to observe the criterion on the chosen candidate. However, there are many other practical cases where this is not a problem, e.g., choosing consumer products.

[7] We use upper case letters to denote random variables, e.g., Y_e, and lower case letters to designate specific values, e.g., y_e. As exceptions to this practice, we use lower case Greek letters to denote random error variables, e.g., ε_e as well as parameters, e.g., $\beta_{e,j}$.

where Y_e represents the criterion (subsequent job performance of candidates) and Y_s is the judgment of the criterion 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 each other and of the X's.

Insert Figure 2 about here

Assuming linearity, 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, i.e., the matches between $\beta_{s,j}$ and $\beta_{e,j}$ for all j = 1,…k. Moreover, the correlation between criterion and judgment, $\rho_{Y_eY_s}$ – the "achievement" index – can be expressed (Tucker, 1964) by

$$\rho_{Y_eY_s} = \rho_{\hat{Y}_e\hat{Y}_s} R_e R_s + \rho_{\varepsilon_e\varepsilon_s}\sqrt{(1-R_e^2)(1-R_s^2)} \quad (3)$$

where $\rho_{\hat{Y}_e\hat{Y}_s}$ (the "matching" index also known as G) is the correlation between the predictions of both models, i.e., between $\sum_{j=1}^{k}\beta_{e,j}X_j$ and $\sum_{j=1}^{k}\beta_{s,j}X_j$; R_e and R_s are, respectively, the multiple correlations of the models of the environment and the judge, and capture, on the one hand, environmental predictability (R_e), and on the other hand, the consistency with which the judge executes the decision rule (R_s). Assuming that the error terms of the two models are independent, i.e., $\rho_{\varepsilon_e\varepsilon_s} = 0$, achievement is simply a multiplicative function of three terms: matching, environmental predictability, and response consistency, and neatly captures the effects of both cognitive and task variables on observed performance or achievement.

Given the above lens model framework, we now develop the probabilities that our models will make correct predictions within a given population or environment.

As will be seen, these probabilities reflect the covariance structure of the cues used as well as those between the criterion and the cues. It is these covariances that describe the inferential environment in which judgments are made. At the same time, we also develop equations for showing the effects of different levels of errors.

The SV model. The lens model – and the lens model equation (3) – have been used extensively to illuminate many issues in judgmental research (Brehmer & Joyce, 1988; Cooksey, 1996). However, here we ask a different question. Imagine that the judge does not decide by using a linear combination rule, but instead simply chooses the candidate who is better on a single variable, X_1, (years of experience, for example). Thus, the decision rule is to choose the candidate for whom X_1 is larger, e.g., 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 or population, 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 (2005b). 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.[8] Denote the correlation by the parameter $\rho_{Y_e X_1}$, ($\rho_{Y_e X_1} > 0$). Given these facts, it is possible to represent Y_{ea} and Y_{eb} by the equations:

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

and $$Y_{eb} = \rho_{Y_e X_1} X_{1b} + v_{eb} \qquad (5)$$

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

[8] We consider the implications of our normality assumption in the Discussion.

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 $M_{f\_SV} = \begin{pmatrix} 2 & 2\rho_{Y_e X_1} \\ 2\rho_{Y_e X_1} & 2 \end{pmatrix}$, and means of 0. Thus the probability of correctly selecting A over B can be written as

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

where $f(d) = \frac{\left|M_f^{-1}\right|^{1/2}}{2\pi} e^{-\frac{1}{2} d' M_f^{-1} d}$ with $d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$.

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, since both its components are equal, can be simplified as

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

The LC model. Following the same rationale, we can also determine the probability that using a linear combination of cues will result in a correct choice. That is, proceeding in exactly the same manner as above, one can express Y_{ea} and Y_{eb} as functions of Y_{sa} and Y_{sb}, define appropriate error terms, ω_a and ω_b, and substitute, respectively, $\rho_{Y_e Y_s}$ for $\rho_{Y_e X_1}$, and Y_{sa} and Y_{sb} for X_{1a} and X_{1b}. Thus, one can show that $2P\{(Y_{ea} > Y_{eb}) \cap (Y_{sa} > Y_{sb})\}$ can also be found through expression (7), with

$F(d)$ 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 $M_{f\_LC} = \begin{pmatrix} 2 & 2\rho_{Y_eY_s} \\ 2\rho_{Y_eY_s} & 2 \end{pmatrix}$.

The EW model. EW is, of course, a special case of LC. Define $d_2 = \overline{X}_a - \overline{X}_b$, where $\overline{X}_a = \frac{1}{k}\sum_{j=1}^{k} X_{ja}$ and $\overline{X}_b = \frac{1}{k}\sum_{j=1}^{k} X_{jb}$, and note that d_2 is a normal variable with a mean of 0.[9] Thus, the expected accuracy of EW can be defined by equation (7) taking into consideration that the appropriate variance/covariance matrix is $M_{f\_EW} = \begin{pmatrix} 2 & 2\rho_{Y_e\overline{X}}\sigma_{\overline{X}} \\ 2\rho_{Y_e\overline{X}}\sigma_{\overline{X}} & 2\sigma_{\overline{X}}^2 \end{pmatrix}$.[10]

The analogous expressions for the CONF and TTB models are presented in Appendix A.

Loss functions. Equation (7) as well as its analogs 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 correct scores that the models achieve in choosing between two alternatives. As such, this measure is equivalent to a 0/1 loss function which does not distinguish between small and large errors. To overcome this deficiency, we 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, &

---

[9] The variable d_1 for EW is the same as for LC: $d_1 = Y_{ea} - Y_{eb}$.

[10] Note that from equation (3) it follows that $\rho_{Y_e\overline{X}} = \rho_{\hat{Y}_e\overline{X}} R_e$ (assuming $\rho_{\varepsilon_e\varepsilon_s} = 0$).

Marquis, 1991). We note that loss occurs when (1) $X_{1a} > X_{1b}$ but $Y_{ea} < Y_{eb}$, and (2) $X_{1a} < X_{1b}$ but $Y_{ea} > Y_{eb}$. Capitalizing on symmetry, the expected loss (EL) associated with the population can therefore be written as

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

where $L = \alpha(Y_{eb} - Y_{ea})^2$. In other words, the expected loss is proportional to the squared difference between Y_{eb} and Y_{ea} weighted by the probability that $Y_{ea} < Y_{eb}$ and $X_{1a} > X_{1b}$. 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:

$$EL_{SV} = 2\alpha(Y_{eb} - Y_{ea})^2 P\{(Y_{ea} < Y_{eb}) \cap (X_{1a} > X_{1b})\} =$$

$$2\alpha \int_{-\infty}^{0} \int_{0}^{\infty} d_1^2 f(d)\ dd \qquad (9)$$

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

Insert Table 1 about here

In Table 1, 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. (We manipulate this factor below.)

Exploring effects of different environments

We first construct and simulate several task environments and demonstrate how our theoretical analyses can be used to make predictions for all of our models in terms of both expected percentage correct predictions 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 thus explicitly explore this issue using simulation. To make the link from theory to empirical phenomena, we report data from a meta-analysis of lens model studies that we use to compare the judgmental performance of theoretical heuristics with that of people using LC models.

Constructed and simulated environments. To demonstrate our approach, we constructed several sets of different three-cue environments using the model implicit in equation (1), i.e.,

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

Our approach was to vary systematically two factors: (1) the weights given to the variables as captured by the distribution of cue validities; (2) the level of average inter-cue correlation. As a consequence, we obtain 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 (due to 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 models[11] (see equations above) with one exception. This is the

[11] For the TTB model, we defined a threshold of 0.50 (with standardized variables) to decide whether a variable discriminated between two alternatives. Whereas the choice of 0.50 was subjective,

LC model which requires specification of $\rho_{Y_e Y_s}$, that is, the "achievement index" or the correlation between the criterion and the person's responses. However, given the lens model equation – see equation (3) above – we know that

$$\rho_{Y_e Y_s} = \rho_{\hat{Y}_e \hat{Y}_s} R_e R_s \qquad (10)$$

where R_e captures the predictability of the environment and $\rho_{\hat{Y}_e \hat{Y}_s} R_s$ the extent to which the person's judgment ability meets the demands of the task, i.e., the product of "matching" and "consistency." [12] Lindell (1976) referred to $\rho_{\hat{Y}_e \hat{Y}_s} R_s$ as "performance" because this part of achievement can be considered separately from task predictability or R_e. We prefer to 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 and consistency of execution.[13] 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 a linear combination of cues with differing levels of "knowledge" (matching of weights) and consistency in execution of their knowledge?

For example, from a psychological perspective an interesting comparison is the point where the use of an LC strategy is equaled by that of a single variable (SV). This occurs when the validity of SV equals that of the person using LC, that is, when $\rho_{Y_e X_1} = \rho_{Y_e Y_s} = caR_e$ or when $ca = (\rho_{Y_e X_1} / R_e)$. One way of thinking about this is to see that, from a predictive viewpoint, it captures the point of indifference between making a judgment using all the data (i.e., with LC) and relying on a single cue (SV)

investigation shows quite similar results if this threshold is varied between 0.25 and 0.75. We use the threshold of 0.5 in all further calculations and illustrations.

[12] The assumption made here is that $\rho_{\varepsilon_e \varepsilon_s} = 0$, see equation (3).

[13] Recall that "using" is employed here in an *as if* manner.

such as representativeness (Kahneman & Tversky, 1972) or affect (Slovic et al., 2002).

The first set of environmental parameters that we consider involves four cases (A, B, C, and D) – see Table 2. Here we examine equal and differential cue validities (case A versus the others), low but positive inter-cue correlations (cases A and C), negative inter-cue correlation (case B), and moderately high inter-cue correlation (case D). These parameters imply different levels of environmental predictability (or lack of "noise"), that is R_e, which varies from 0.66 to 0.93. In the right hand column, we show values of ($\rho_{Y_e X_1}/R_e$) which indicate the benchmarks for determining when SV or LC performs better. Specifically, LC performs better than SV when *ca* exceeds ($\rho_{Y_e X_1}/R_e$).

Insert Table 2 and Figures 3, 4, and 5 about here

Figure 3 depicts expected percentage correct predictions of the different models as a function of linear cognitive ability or *ca*. In addition, Figure 3 recognizes the possibility that the decision maker could err in using the SV and TTB models – specifically by failing to order the variables according to their cue validities. This is shown in respect of SV in the four left-hand side panels and for TTB in the four right-hand side panels. Here the lines SVr and TTBr show expected performance if cues are selected or ordered at random and the shaded areas indicate the range of possible performance levels from best (the correct order) to worst (most incorrect order).

A first comment is that, in a relative sense, model performance varies by environments. In case A (equal cue validities and low cue inter-correlation), for example, EW performs best and CONF is also more effective than TTB. SV lags behind. Note that, in this environment, it does not matter whether heuristics identify the correct ordering of cues because each cue has the same validity.

This picture changes when the cue validities differ. In case B (with negative inter-cue correlation), EW is still best, but only slightly, whereas TTB now outperforms CONF. As cue inter-correlation increases, however, differences in model performance decrease – examine cases C and D – and EW no longer has the best performance. As can also be seen, errors in failing to identify the correct ordering of cues can hinder performance in environments B, C, and D.

Second, consider the performance of LC as a function of *ca*. First note that equality between LC and SV occurs, for each of the cases, at the critical points enumerated at the right of Table 2. Thus, for example, LC needs less linear cognitive ability in case A (0.62) to do better than SV than in case C (0.80). Interestingly, in all the environments illustrated, linear cognitive ability has to be quite high before it starts to be competitive with the better heuristics. Indeed, it is only in case B that LC has the best performance and this when linear cognitive ability starts to exceed 0.85.

The simple conclusion from this analysis – which we explore further below – is that unless linear cognitive ability is "high," decision makers are better off using simple heuristics provided that they implement these correctly.

In Figure 4, we show differential performance in terms of expected loss where the exactingness parameter, α, is equal to 1.00. A comparison of Figures 3 and 4 shows the same pattern of results in terms of relative model performance. Once again, we also illustrate the effects of errors in the use of SV and TTB. Figure 5 examines the effects of less exacting losses when $\alpha = 0.30$. Compared to Figure 4, we find the same relative ordering between models but differences in expected loss are much smaller. Indeed, the effect of changing α is to reduce or magnify (as appropriate) expected losses by a constant multiplier (see note 2 to Table 1).

To provide more insight, we constructed four further sets of environments – cases E, F, G, and H – each of which had eight sub-cases (i through viii) as specified in Table 3. In cases E and F, the distribution of cue validities was quite steep (decreasing constantly by one-half) and overall cue validity decreased across sub-cases (i through viii). Cases G and H had a similar design except that the distribution of cue validities was flatter. Cases E and G had low positive cue inter-correlation whereas cases F and H had higher cue-intercorrelation. A consequence of these specifications was a range of environmental predictabilities (R_e) from 0.37/0.39 to 0.85/0.88 across all eight sets of sub-cases.

Insert Table 3 and Figures 6 and 7 about here

Table 3 also documents expected percentage correct and losses (for $\alpha = 1.00$) for all our models including LC which has been calculated using three different values for linear cognitive ability: *ca* = 0.5 for LC1; *ca* = 0.7 for LC2; and *ca*= 0.9 for LC3. The trends in Table 3 are perhaps better viewed by examining Figures 6 and 7 that document percentage correct and expected loss, respectively, of the different models as a function of the validity of the most valid cue, $\rho_{Y_e X_1}$. Since here $\rho_{Y_e X_1}$ is highly correlated with R_e, the horizontal axis of the graphs can also be thought of as capturing "noise" (more, on the left, to less, at the right). As with Figures 3 and 4, we use shaded areas to indicate the ranges of performance that can be achieved by SV (on the left) and TTB (on the right).

Abstracting first from the three LC models, there is a general trend (that could be expected) for differences in model performance to increase as noise or error in the environment decreases. TTB dominates the other models in case E but is, in turn, dominated by SV in the more redundant case F. In case G (where the distribution of cue validities is flatter), EW and TTB are the better performing models, and EW does better than TTB when $\rho_{Y_e X_1} < 0.50$. In case H (involving greater redundancy), SV is, once again, one of the better models. CONF generally tracks EW closely but is consistently inferior to it. The difference between looking at percentage correct (Figure 6) and expected loss with $\alpha = 1.00$ (Figure 7) is that differences between models are easier to observe with the latter.

In terms of linear cognitive ability, it is clear (and unsurprising) that more is better than less. Interestingly, however, as the environment becomes more predictable the effectiveness of the LC models drops off relative to the simpler heuristics. (This can also be seen by considering the $\rho_{Y_e X_1}/R_e$ column in Table 3.) In the environments examined here, the best LC model (with $ca = 0.9$) is always outperformed by one of the other heuristics when $\rho_{Y_e X_1} > 0.60$.

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, since calculating this joint probability is quite complicated in some cases, we actually simulated results based on 5,000 trials for each environment.

Table 4 specifies the parameters of the environments we considered, the percentage correct predictions for each model in each environment,[14] and the

[14] We had also calculated the theoretical probabilities of the simulated percentage correct predictions. Given the large sample sizes (5,000), theoretical and simulated results were almost identical.

probabilities that models would make the same decisions. As can be seen, there are two sets of environments, I and J, each with eight sub-sets (i through viii). Set I has low cue inter-correlation; set J has moderate to high cue inter-correlation. Within each set, we vary predictability (R_e) from high to low.

Insert Table 4 about here

We make three remarks. First, whereas there is considerable variation in percentage correct predictions across different levels of predictability, 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 0.70 and 0.92 across all comparisons and is probably higher than one might have imagined a priori. 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, and as would be expected, the effect of increasing cue inter-correlation (or redundancy) is to increase the level of agreement between models. Third, for the environments illustrated here, the CONF and EW models have the highest level of agreement whereas the SV-EW and SV-TTB have the lowest. The latter result is perhaps surprising in that both SV and TTB are so dependent on the most valid cue.

Comparisons with experimental data. Although instructive, 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 (3). Studies therefore had to have a criterion variable and involve the judgments of individuals (as opposed to groups of people).[15] Moreover, we only considered cases where the number of independent variables or cues was greater or equal to two (when there is only one cue, $\rho_{\hat{Y}_e \hat{Y}_s}$ = 1.00 necessarily). In all, we located 77 (mainly) published papers that allowed us to examine judgmental performance across 252 different task environments (i.e., environments that vary by statistical parameters and/or substantive conditions).

In Table 5, we summarize key statistics from the meta-analysis (for full details, see Karelaia & Hogarth, in preparation). 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, since 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 inter-correlation).

Insert Table 5 about here

Briefly, we find no differences in performance between participants who are experts or novices (the latter, however, are assessed after learning) nor between laboratory and field studies. Holding the predictability of the environment constant (i.e., R_e), performance (both r_a and LC accuracy) is somewhat better with fewer cues,

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

and with equal as opposed to differential weighting functions. Parenthetically, in characterizing the latter, we classify functions as non-compensatory if, when cue validities are ordered in magnitude, the validity of each cue exceeds the sum of those smaller than it (cf., Martignon & Hoffrage, 1999; 2002). We define all other functions as compensatory except for the special case of equal-weighting.

Overall, the LC accuracy reported in the right hand column of Table 5 is about 70%. In interpreting this figure, it is important to bear in mind that it is derived from an estimate of linear cognitive ability (*ca* or GR_s) of 0.66 and that this figure is a mean estimate across individual studies each of which is described by the mean of individual data. Table 5 obscures individual variation.

To capture the differences in performance between LC and the heuristic models, one needs specific information on the statistical properties of tasks (essentially the covariation matrix 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 the measure of achievement into one of performance in binary choice using the methods described above, that is, by assessing the performance of LC with different levels of linear cognitive ability, *ca*. Thus, to measure the effectiveness of LC relative to any heuristic in a particular environment, we considered the difference in expected predictive ability between LC based on the mean *ca* observed in the environment and that of the heuristic. In other words, we ask how well the *average* performance levels of humans using LC compare to those of heuristics.

In Table 6, 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 we 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, various forms of feedback, presentation of information, and so on).

Insert Table 6 about here

The upper panel of Table 6 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 non-compensatory functions, respectively. This captures the predictability of the environments (81% for equal weighting and compensatory and 82% for non-compensatory). These environments are also marked by little redundancy. Over 80% have mean inter-cue correlations of 0.00. In the body of the table, we present performance in terms of percentage correct for LC – based on *mean* cognitive ability 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. Thus, one way of interpreting the LC column is as the performance that would have been achieved in binary choice by the mean participant in each study (in terms of judgmental ability).

As would be expected, the EW strategy performs best in equal weighting environments (80%) and the TTB strategy best in the non-compensatory environments (77%). Interestingly, in these compensatory environments, it is the EW model that performs best (77%). The mean LC model never has the best performance. Compared to 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 in using heuristics. 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 6, we present the data based on analyzing studies with two cues where, once again, most environments involve orthogonal cues (73%) – 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 non-compensatory environments although it is bettered here by the SV model (just).[16]

Since most published studies do not report individual data, it is difficult to assess the importance of individual variation in performance in particular tasks and, specifically, how individual LC performance compares with heuristics in such tasks. Two papers involving two-cues did report the necessary data (Steinmann & Doherty, 1972; York et al., 1987). Table 7 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 percentage of participants that have better performance with LC than with particular heuristics. (Note the three tasks reported by York et al., 1987 have identical statistical characteristics but involved different substantive manipulations of information).

[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.

Clearly, one cannot generalize from the four environments presented in Table 7. However, it is of interest to note that, first, the ranges of individual LC performances are quite large (24% to 31%), and second, a limited number of participants can have better performance with LC than with the heuristics.

Insert Table 7 about here

Summary. At a theoretical level, we have shown that the performance of heuristic rules is affected by several factors: the type of weighting function (i.e., how the environment weights cues); cue redundancy or inter-correlation; the predictability of the environment; and loss functions. The weighting function determines which heuristic is best suited to specific tasks and this depends on how its characteristics match that of the tasks confronted. For example, EW is better in equal weighting/compensatory environments and TTB and SV in non-compensatory environments. The effect of cue redundancy is generally to reduce differences in the relative predictive abilities of the heuristics. As environments become more predictable, all models perform better but differences between models also increase. Finally, the effect of loss functions is to accentuate or dampen differences between evaluations of model predictions.

We also used simulation to investigate the extent to which models agree with each other. At one level, all the models we investigated were “sensible” and used valid information. As such, it should not be surprising that they exhibited much agreement. The extent of the agreement, however, was surprising. Even when the predictability of the environment varied greatly, the level of agreement between particular models hardly changed (cf., Table 4). From a predictive viewpoint, this might be thought comforting. But 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 – and thus the importance of identifying when different heuristics are more effective.

The differential impact of environmental factors is illustrated quantitatively in Table 8 which reports the results of regressing performance of the heuristics (percentage correct) on environmental factors: type of weighting function (represented by dummy variables), redundancy (cue inter-correlation), and predictability (R_e). This is done for the 52 populations specified in Tables 2, 3, and 4. Results show the importance of non-compensatory environments and redundancy on SV (positive) and EW and CONF (both negative). Interestingly, for the conditions examined here, the performance of TTB is not affected by these factors thereby suggesting a heuristic that is robust to environmental variations (for further analysis of this issue, see Baucells, Carrasco, & Hogarth, 2006). Finally, all models benefit from greater predictability.

Insert Table 8 about here

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

Our theoretical analyses suggested that *ca* needed to be larger than about 0.7 for LC models to perform better than heuristics. Across the 252 task environments of the meta-analysis we estimated *ca* to be 0.66. However, this is a mean and does not take account of differences in task environments. For those environments where precise predictions could be made, LC models based on mean *ca* estimates performed at a level inferior to the best heuristics but equal to or better than heuristics executed with error. Unfortunately, the data did not allow us to make a thorough investigation of individual variation in *ca* values. However, to the extent that we could do this, only a minority of individuals appeared capable of outperforming heuristics using LC.

General discussion

Our goal has been to show 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 statistical properties. Given the inherent uncertainty in inference, this approach seems eminently sensible (cf., Brunswik, 1952). We now consider implications that are, first, psychological, second, normative, and third, methodological in nature. We also outline extensions for further work.

Psychological implications. All of the models (heuristics) we have examined can be thought of representing "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., Hoffman, 1960; Einhorn et al., 1979) and – apart from predictive tests – we believe the answer lies in assessing logical consistency between the assumptions of models and the information processing operations actually performed by humans.

Consider, for example, the SV (the simplest) and the LC (arguably the most complex) 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 would need to show is that mental operations took place that led to outcomes consistent with the operations. Nor do we 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.

From a psychological viewpoint, therefore, the claim that the different models capture actual processes is made at a level of analysis that represents mental operations in an "as if" manner. Moreover, by defining the statistical properties of task environments, we show at a theoretical level how characteristics of models and tasks result in different levels of performance. This is an important contribution because it provides the basis for developing an environmental theory of judgmental performance (cf., Brunswik, 1952; Simon, 1956).

The environment, however, is not captured by statistical properties alone since context can be important. Within our framework, contextual effects would be reflected in how people use heuristics. Consider, for example, what happens when cue variables are inappropriately labeled. Within LC models, this would be captured by reductions in linear cognitive ability (*ca*) because people give less appropriate weights to the variables. With the TTB model, it could result in using cues 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. Since it is people who are differentially susceptible to contextual effects, we believe this makes sense.

One claim we do make is that the range of models we considered covers the types of heuristics that have been discussed in the literature as well, of course, as the linear model. Thus, the SV model captures precisely what happens when people make decisions based on a single cue such as representativeness (Kahneman & Tversky, 1972), 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 can not 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. From our perspective, it is understandable that some researchers see the "glass as half-empty" while 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 TTB, 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.

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 both G and R_s are approximately 0.80 (Table 5). Moreover, only 11% of GR_s values exceed 0.90. That is, the meta-analysis reveals much error in both knowledge and execution. Note also that although G and R_s are positively correlated, 0.43 ($p < .001$), neither G nor R_s are correlated with the predictability of the environment (R_e) – 0.03 for G and 0.09 for R_s. In other words, there is a trend for people to be more consistent in executing strategies when these are more valid. However, there is no relation between how predictable an environment is and people's judgmental strategies other than a kind of probability matching result where, overall, mean R_e and R_s are approximately equal.

Given the difficulty of executing the LC model well, it is of interest to speculate when people can rely on this kind of process. We suspect that many models of this type – or "as if" versions – are used when judgmental processes have been automated (or become "tacit," Hogarth, 2001) such that people do not need to think about executing trade-offs. Imagine, for example, basic processes such as perception or situations where past practice has been sufficient to hone a person's skills. These include the judgments that most of us 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 et al., 2005).

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. This issue emphasizes the need to understand the natural ecology of decision making tasks (Dhami, Hertwig, & Hoffrage, 2004).

Normative implications. Our work has many normative implications in that it spells out the conditions under which different heuristics are effective. Moreover, the fact that this is achieved analytically – instead of through simulation – represents an advance over current practice (see also Hogarth & Karelaia, in press b).

An interesting normative implication relates to the trade-offs in different types of error when using heuristics or models. As noted above, one way of characterizing our empirical analysis is to say that judgmental performance using the cognitively demanding LC models is roughly equal to that of using heuristics with error, that is, of SVr and TTBr. However, is there a relation between linear cognitive ability (*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 how the individual acquired the necessary knowledge. Basically – at one extreme – if all cues are approximately equally valid, EW should be used explicitly. Similarly – at the other extreme – when facing a non-compensatory 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 linear cognitive ability (*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? Based on our meta-analysis, we expect that a minority of persons can meet these conditions but that much also depend on the nature of the task and the individual's predictive 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. This, in turn, raises psychological issues of how people acquire such knowledge or are helped to do so. 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. 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 (cf., Einhorn, 1986).

Methodological implications. Our work involves methodological innovations. Not only have we developed analytical tools for problems that frequently use simulation, we have also 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.

Several issues suggested further work. First, in this paper, we have limited ourselves to a binary choice paradigm involving three cues. This can be extended in two ways: first, to consider more alternatives, and second, more cues. Our previous work (Hogarth & Karelaia, in press a, b), 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 inter-cue correlation.

Second, 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 if distributions were skewed and/or had fatter tails than the normal distribution? Further interesting complications could involve effects where models have correlated error terms.

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

Concluding comments. As noted at the outset of this paper, our goal has not been to "unify" different traditions of judgmental research. However, we have developed a framework in which to compare results. Thus, we have been able to make direct comparisons between research in the long-standing lens model tradition with the more recent work on heuristic decision making. Central to our approach has been the need to specify and model characteristics of task environments for it is this that determines which and why particular heuristics are more or less successful. It also provides guidance as to the level of expertise needed to use the more demanding LC models. At the same time, we emphasize the need for knowledge – or maps – to know when to use specific heuristics. How people develop such maps is key to understanding much judgmental activity.

References

Anderson, N. H. (1981). *Foundations of information integration theory*. New York, NY: Academic Press.

Armelius, B-A, & Armelius, K. (1974). The use of redundancy in multiple-cue judgments: Data from a suppressor-variable task. *American Journal of Psychology*, 87 (3), 385-392.

Ashton, R. H. (1981). A descriptive study of information evaluation. *Journal of Accounting Research*, 19 (1), 42-61.

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

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.

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 (5), 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 (4), 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 (2), 277-293.

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: The University of Chicago Press.

Camerer, C. F. (1981). General conditions for the success of bootstrapping models. *Organizational Behavior and Human Performance*, 27, 411-422.

Chaiken, S., & Trope, Y. (1999). (Eds.) *Dual-process theories in social psychology*. New York: Guilford Press.

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, NY: Academic Press.

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

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

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). 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 (1), 20-26.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, 130(6), 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 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 (3), 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.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49,* 709-724.

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

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, 103 (3), 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., and 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 (1), 75-90.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, 62 (4), 255-262.

Hammond, K. R. (1990). Functionalism and illusionism: Can integration be usefully achieved. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 227-261). Chicago, IL: The University of Chicago Press.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford: Oxford University Press.

Hammond, K. R., & Stewart, T. R. (Eds.). (2001). *The essential Brunswik: Beginnings, explications, applications*. New York, NY: 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., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review,* 71, 438–456.

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 (2), 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, UK: Wiley.

Hogarth, R. M. (2001). *Educating intuition*. Chicago, IL: The University of Chicago Press.

Hogarth, R. M., & Karelaia, N. (2005a). 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*, 51 (12), 1860-1872.

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

Hogarth, R. M., & Karelaia, N. (in press, a). "Take-the-Best" and other simple strategies: Why and when they work "well" in binary choice. *Theory and Decision*.

Hogarth, R. M., & Karelaia, N. (in press, b). Regions of rationality: Maps for bounded agents. *Decision Analysis*.

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(4), 734-752.

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., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430-454.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103 (3), 582-591.

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

Karelaia, N. (2006). Thirst for confirmation in multi-attribute choice: Does search for consistency impair decision performance? *Organizational Behavior and Human Decision Processes*. 100 (1), 128-143.

Karelaia, N., & Hogarth, R. M. (in preparation). *Determinants of linear judgment: a meta-analysis of lens studies*. Lausanne and Barcelona.

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

Keeney, R., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value trade-offs*. New York, NY: Wiley.

Kessler, L., & Ashton, R. H. (1981). Feedback and prediction achievement in financial analysis. *Journal of Accounting Research*, 19 (1), 146-162.

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(2), 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 (6), 739-745.

Louvière, J. J. (1988). *Analyzing decision making: Metric conjoint analysis*. Thousand Oaks, CA: Sage Publications.

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 and the ABC Research Group. *Simple heuristics that make us smart* (pp. 119-140). New York, NY: Oxford University Press.

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

Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

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(1), 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.

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 and the ABC Research Group. *Simple heuristics that make us smart* (pp. 141-167). New York: Oxford University Press.

Rieskamp, J., & Hoffrage, U. (2002). *The use of simple heuristics: Inferences and preferences under time pressure*. Manuscript, Max Planck Institute for Human Development, Berlin.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental psychology* (Vol 10): New York: Academic Press.

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.

Russo, J. E., & Schoemaker, P. J. H. (2002). *Winning decisions: Getting it right the first time.* New York, NY: Doubleday.

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

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.

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, NY: Cambridge University Press.

Stanovich. K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, 127, 161-188.

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. and 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 (2), 203-211.

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

Todd, P. M., & Gigerenzer, G. (2000). Précis of *Simple heuristics that make us smart*. *Behavioral and Brain Sciences*, 23, 727-780.

Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hursch, Hammond and Hursch and by Hammond, Hursch 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). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-131.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, 83, 213-217.

Wallace, H. A. (1923). What is in the corn judge's mind? *Journal of the American Society of Agronomy*, 15, 300-304.

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 they own self be true: Finding the utility of cognitive information feedback. *Journal of Behavioral Decision Making*, 18, 319-341.

Table 1 – Key formulas for three models: SV, LC, and EW

Model	variance-covariance matrix (M_f)
Single variable (SV)	$\begin{pmatrix} 2 & 2\rho_{Y_e X_1} \\ 2\rho_{Y_e X_1} & 2 \end{pmatrix}$
Linear Combination (LC)	$\begin{pmatrix} 2 & 2\rho_{Y_e Y_s} \\ 2\rho_{Y_e Y_s} & 2 \end{pmatrix}$
Equal weights (EW)	$\begin{pmatrix} 2 & 2\rho_{Y_e \overline{X}}\sigma_{\overline{X}} \\ 2\rho_{Y_e \overline{X}}\sigma_{\overline{X}} & 2\sigma_{\overline{X}}^2 \end{pmatrix}$

Notes:

1. *The expected accuracy* of models is estimated as the probability of correctly selecting A over B, and is found as:

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

where $f(d) = \frac{\left|M_f^{-1}\right|^{1/2}}{2\pi} e^{-\frac{1}{2}d' M_f^{-1} d}$ with $d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$.

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. $\rho_{Y_e \overline{X}} = \overline{\rho}_{Y_e X} \sqrt{\dfrac{k}{1 + (k-1)\overline{\rho}_{X_i X_j}}}$, where k = number of X variables, $\overline{\rho}_{Y_e X}$ = average correlation between Y and the X's, and $\overline{\rho}_{X_i X_j}$ = average intercorrelations amongst the X's.

5. $\sigma_{\overline{X}} = \sqrt{\dfrac{1}{k}\left(1 + (k-1)\overline{\rho}_{X_i X_j}\right)}$

Table 2 -- Environmental parameters: Cases A, B, C, and D

	Cue validities			Cue inter-correlations			R_e	$\rho_{Y_e X_1}/R_e$
	$\rho_{Y_e X_1}$	$\rho_{Y_e X_2}$	$\rho_{Y_e X_3}$	$\rho_{X_1 X_2}$	$\rho_{X_1 X_3}$	$\rho_{X_2 X_3}$		
Case A	0.5	0.5	0.5	0.1	0.1	0.1	0.81	0.62
Case B	0.6	0.4	0.3	-0.4	0.1	0.1	0.93	0.64
Case C	0.6	0.4	0.3	0.1	0.1	0.1	0.75	0.80
Case D	0.6	0.4	0.3	0.5	0.5	0.5	0.66	0.91

Table 3 -- Environmental parameters: Cases E, F, G, and H; and performance (expected percentage correct and expected losses)

	Cue validities			Cue inter-correlations			R_e	$\rho_{Y_eX_1}/R_e$	Percentage correct							Loss (α = 1.00)						
	$\rho_{Y_eX_1}$	$\rho_{Y_eX_2}$	$\rho_{Y_eX_3}$	$\rho_{X_1X_2}$	$\rho_{X_1X_3}$	$\rho_{X_2X_3}$			LC1*	LC2*	LC3*	SV	EW	TTB	CONF	LC1*	LC2*	LC3*	SV	EW	TTB	CONF
Case E																						
i	0.8	0.4	0.2				0.88	0.91	64	71	79	80	76	**82**	74	0.5	0.3	0.1	0.1	0.2	**0.1**	0.2
ii	0.7	0.35	0.175				0.78	0.90	63	68	75	75	72	**77**	70	0.5	0.3	0.2	0.2	0.2	**0.2**	0.3
iii	0.6	0.3	0.15				0.69	0.87	61	66	71	70	69	**72**	67	0.6	0.4	0.3	0.3	0.3	**0.3**	0.4
iv	0.5	0.25	0.125	all equal to 0.1			0.60	0.83	60	64	68	67	65	**68**	64	0.6	0.5	0.3	0.4	0.4	**0.4**	0.5
v	0.4	0.2	0.1				0.52	0.76	58	62	66	63	62	**64**	61	0.7	0.5	0.4	0.5	0.5	**0.5**	0.6
vi	0.3	0.15	0.075				0.45	0.66	57	60	63	60	59	**61**	58	0.7	0.6	0.5	0.6	0.7	**0.6**	0.7
vii	0.2	0.1	0.05				0.41	0.49	56	59	62	56	56	**57**	55	0.7	0.6	0.6	0.8	0.8	**0.7**	0.8
viii	0.1	0.05	0.025				0.37	0.27	56	58	61	53	53	**54**	53	0.8	0.7	0.6	0.9	0.9	**0.9**	0.9
Case F																						
i							0.85	0.94	64	70	78	**80**	69	76	69	0.5	0.3	0.1	**0.1**	0.3	0.2	0.4
ii							0.76	0.92	62	68	74	**75**	67	72	66	0.5	0.4	0.2	**0.2**	0.4	0.2	0.4
iii							0.67	0.89	61	66	71	**70**	64	68	64	0.6	0.4	0.3	**0.3**	0.5	0.3	0.5
iv	same as in Case E			all equal to 0.5			0.59	0.85	59	63	68	**67**	62	65	61	0.6	0.5	0.4	**0.4**	0.6	0.4	0.6
v							0.51	0.78	58	62	65	**63**	59	62	59	0.7	0.6	0.4	**0.5**	0.6	0.5	0.7
vi							0.45	0.67	57	60	63	**60**	57	59	57	0.7	0.6	0.5	**0.6**	0.7	0.7	0.7
vii							0.40	0.49	56	59	62	**56**	55	**56**	54	0.7	0.6	0.6	**0.8**	0.8	0.8	0.8
viii							0.37	0.27	56	58	61	**53**	52	**53**	52	0.8	0.7	0.6	**0.9**	0.9	0.9	0.9
Case G																						
i	0.8	0.4	0.2				0.88	0.91	64	71	79	80	76	**82**	74	0.5	0.3	0.1	0.1	0.2	**0.1**	0.2
ii	0.7	0.4	0.2				0.80	0.88	63	69	76	75	74	**78**	71	0.5	0.3	0.2	0.2	0.2	**0.1**	0.3
iii	0.6	0.4	0.2				0.73	0.83	62	67	73	70	72	**73**	69	0.6	0.4	0.2	0.3	0.3	**0.2**	0.3
iv	0.5	0.4	0.2	all equal to 0.1			0.66	0.75	61	65	70	67	**70**	**70**	67	0.6	0.4	0.3	0.4	**0.3**	0.3	0.4
v	0.4	0.4	0.2				0.61	0.65	60	64	69	63	**68**	66	66	0.6	0.5	0.3	0.5	**0.4**	0.4	0.4
vi	0.3	0.3	0.2				0.52	0.57	58	62	66	60	**64**	62	62	0.7	0.5	0.4	0.6	**0.5**	0.6	0.5
vii	0.2	0.2	0.2				0.45	0.44	57	60	63	56	**60**	58	59	0.7	0.6	0.5	0.8	**0.6**	0.7	0.6
viii	0.1	0.1	0.1				0.39	0.26	56	59	61	53	**55**	54	**55**	0.8	0.7	0.6	0.9	**0.8**	0.8	0.8
Case H																						
i							0.85	0.94	64	70	78	**80**	69	76	69	0.5	0.3	0.1	**0.1**	0.3	0.2	0.4
ii							0.76	0.93	62	68	74	**75**	68	73	67	0.5	0.4	0.2	**0.2**	0.4	0.2	0.4
iii							0.67	0.89	61	66	71	**70**	66	**70**	66	0.6	0.4	0.3	**0.3**	0.4	0.3	0.4
iv	same as in Case G			all equal to 0.5			0.60	0.83	60	64	68	**67**	65	**67**	64	0.6	0.5	0.3	0.4	0.5	**0.4**	0.5
v							0.54	0.74	59	62	66	63	63	**64**	63	0.7	0.5	0.4	0.5	0.5	**0.5**	0.5
vi							0.47	0.64	58	61	64	60	**61**	**61**	60	0.7	0.6	0.5	0.6	**0.6**	**0.6**	0.6
vii							0.41	0.48	57	59	62	56	**58**	57	**58**	0.7	0.6	0.5	0.8	**0.7**	0.7	0.7
viii							0.37	0.27	56	58	61	53	**54**	**54**	**54**	0.8	0.7	0.6	0.9	**0.8**	0.9	0.9

* For LC1, c = 0.5; for LC2, c = 0.7; for LC3, c = 0.9.

Notes: The performance of the best heuristic in each environment is highligted with **bold** 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.

Table 4 -- Rates of agreement between heuristic strategies for different environments*

	Cue validities			Cue inter-correlations			R_e	Percentage correct				Rates of agreement					
	$\rho_{Y_eX_1}$	$\rho_{Y_eX_2}$	$\rho_{Y_eX_3}$	$\rho_{X_1X_2}$	$\rho_{X_1X_3}$	$\rho_{X_2X_3}$		SV	EW	TTB	CONF	SV-EW	SV-CONF	SV-TTB	CONF-EW	TTB-EW	CONF-TTB
Case I																	
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		all equal to 0.1		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
							Means	65	67	68	65	0.72	0.77	0.72	0.86	0.79	0.77
Case J																	
i							0.90	80	73	79	71	0.81	0.83	0.81	0.91	0.88	0.85
ii							0.78	74	70	74	68	0.81	0.84	0.81	0.91	0.88	0.85
iii							0.67	71	67	70	65	0.81	0.83	0.81	0.91	0.88	0.84
iv		same as in Case I			all equal to 0.5		0.58	67	64	66	63	0.80	0.83	0.80	0.91	0.88	0.85
v							0.50	64	61	63	60	0.80	0.83	0.80	0.91	0.87	0.84
vi							0.44	59	59	60	59	0.81	0.84	0.81	0.91	0.87	0.84
vii							0.42	58	59	58	59	0.80	0.83	0.80	0.91	0.88	0.85
viii							0.38	53	54	53	53	0.80	0.83	0.80	0.92	0.88	0.84
							Means	66	63	65	62	0.81	0.83	0.81	0.91	0.88	0.85
							Overall means	66	65	66	64	0.76	0.80	0.76	0.88	0.84	0.81

* Results are from simulations with 5,000 trials for each environment.

Table 5: Description of studies in lens model meta-analysis

	No. of studies	Average number of: judges	judgments	r_a*	Mean lens model statistics G*	Re	Rs	C*	GRs	LC accuracy (%)
Characteristics of tasks										
Participants:										
Experts	59	23	102	0.57	0.80	0.82	0.79	0.06	0.65	71
Novices	192	19	92	0.55	0.81	0.79	0.81	0.06	0.66	70
Unclassified	4									
Type of study:										
Laboratory	200	21	93	0.56	0.82	0.80	0.79	0.04	0.67	70
Field	51	15	95	0.55	0.76	0.77	0.85	0.11	0.66	70
Unclassified	4									
Number of cues:										
2	67	26	58	0.63	0.88	0.80	0.79	0.07	0.70	73
3	84	19	98	0.55	0.87	0.81	0.80	0.00	0.72	70
> 3	96	16	111	0.51	0.71	0.79	0.81	0.08	0.58	68
Unclassified	8									
Type of weighting function:										
Equal weighting	40	31	82	0.66	0.90	0.82	0.80	0.02	0.74	75
Compensatory	84	16	102	0.58	0.84	0.81	0.83	0.04	0.70	71
Non-compensatory	50	23	41	0.50	0.79	0.84	0.72	0.04	0.60	67
Unclassified	81									
Cue redundancy:**										
None	92	22	56	0.61	0.88	0.83	0.80	0.03	0.72	72
Low-medium	79	19	98	0.53	0.79	0.79	0.83	0.03	0.66	68
High	26	25	105	0.54	0.77	0.75	0.80	0.10	0.64	69
Unclassified	58									

Notes:
*These statistics correspond to the sample estimates of the elements of the lens model equation presented in the text -- equation (3).
(*ra* 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, $\rho_{\varepsilon_e \varepsilon_s}$).
** We define redundancy by the level of average inter-cue correlation. None implies the absolute value of average intercorrelation of 0; low-medium -- the absolute value of <=0.4 (also described in text as "low", "moderate", "some"); and high -- the absolute value of >0.4 (also described in text as "a lot", "high").

Table 6: Performance of heuristics and mean LC in 3-cue and 2-cue environments

Weighting function	Maximum possible percentage correct	LC[1]	SV	SVr	EW	CONF	TTB	TTBr	Numbers of environments
		Performance -- Percentage correct using:							
3- cue environments[2]									
Equal weighting	81	72	65	65	**80**	74	71	70	9
Compensatory	81	68	69	64	**77**	72	73	68	19
Non-compensatory	82	67	73	63	74	70	**77**	67	26
								Subtotal	54
2 - cue environments[3]									
Equal weighting	94	79	73	73	**92**	73	80	80	12
Non-compensatory	84	69	**76**	67	73	67	75	70	21
								Subtotal	33
								Total	87

Notes:
1 -- Based on empirically observed mean linear cognitive ability (*ca*).
2 -- Averages calculated on the 54 environments detailed in Appendix C.
3 -- Averages calculated on the 33 environments detailed in Appendix D.
Bold indicates largest percentage correct in each row.

Table 7: Levels of individual performance relative to heuristics

	Number of participants	Statistical properties of tasks				LC performance (% correct)			Percentage of participants with better performance than:					
		R_e	$\rho_{Y_e X_1}$	$\rho_{Y_e X_2}$	$\rho_{X_1 X_2}$	Mean	Max	Min	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 et al. (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

Table 8 -- Regression of model performance (percentage correct) on environmental characteristics for populations in Tables 2, 3, and 4

	SV	EW	TTB	CONF
Regression coefficients				
Intercept	34.1	43.0	36.7	43.1
t - statistic	31.1	46.2	44.3	48.7
Dummy: compensatory	2.0*			
t - statistic	2.5			
Dummy: non-compensatory	3.5	-2.4		-1.7
t - statistic	4.3	-5.3		-3.9
Redundancy	6.3	-6.1		-3.2
t - statistic	4.9	-5.6		-3.0
Predictability (R_e)	45.2	40.5	48.9	36.3
t - statistic	27.0	31.4	37.5	29.6
Adjusted R^2	0.95	0.96	0.97	0.95

Notes:
(1) The regressions are based on 52 observations. The dummy variables for compensatory and non-compensatory weighting functions are expressed relative to equal weighting which is captured within the intercept term.
(2) There are only three levels of redundancy: mean inter-cue correlation of -0.07, 0.1, and 0.5.
(3) Only statistically significant coefficients are shown. All coefficients are significant ($p < .001$) except when marked * for $p < .05$.

Figure 1: Knowledge required to achieve upper limits of model performance

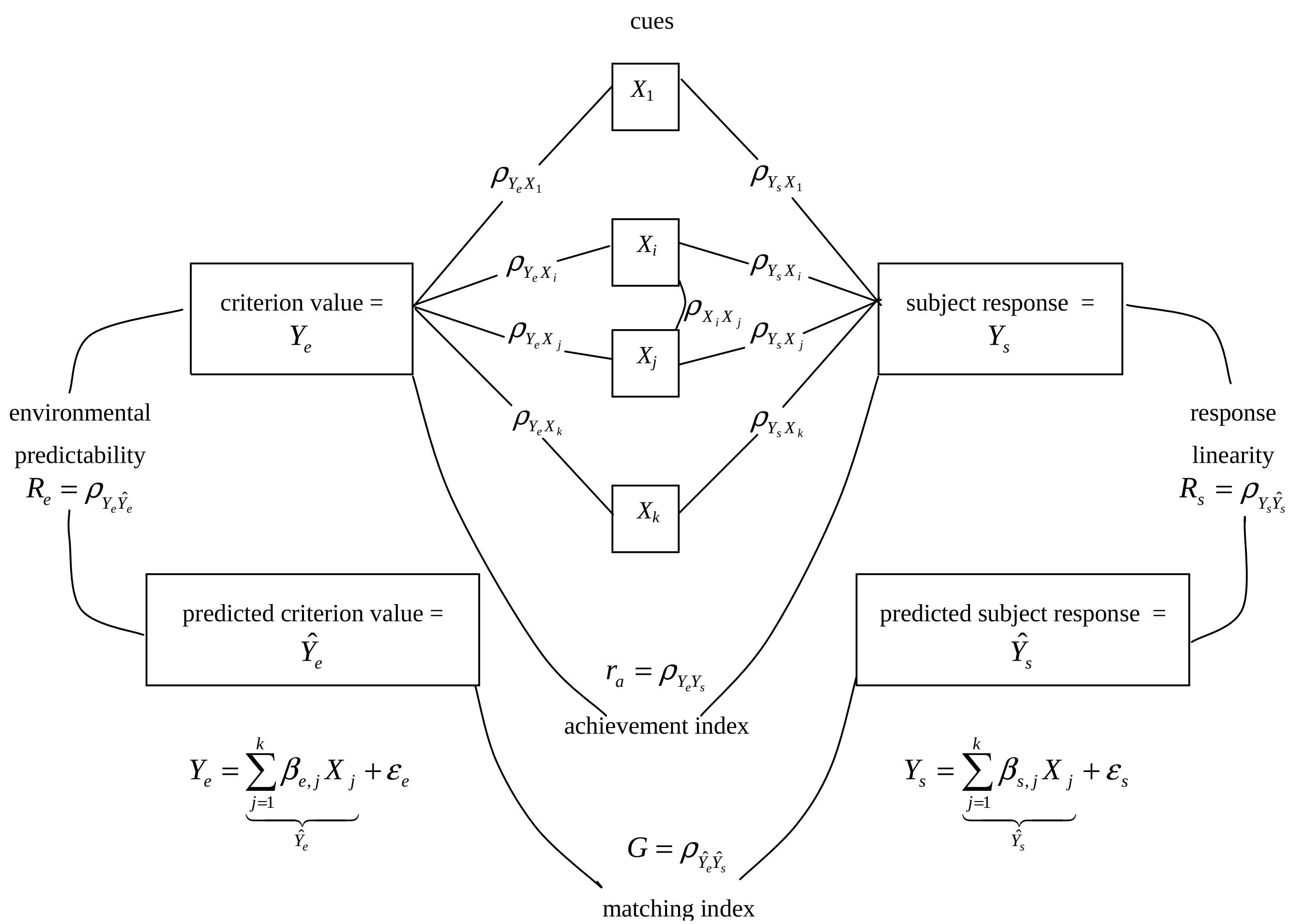
	Values of variables[1]			Weights Ordering			
Model	Cue 1	Cue 2	Cue 3	"Exact"[2]	First[3]	All[3]	None
Linear combination (LC)	Yes	Yes	Yes	Yes			
Equal weighting (EW)	Yes	Yes	Yes				Yes
Take-the-best (TTB)	Yes	Yes/No	Yes/No			Yes	
Single variable (SV)	Yes				Yes		
CONF	Yes	Yes	Yes/No				Yes

[1] Yes = value of cue required; Yes/No = value of cue may be required.

[2] Exact values of cue weights required.

[3] First = most important cue identified; All = rank order of all cues known a priori.

Figure 2: Diagram of lens model.



Figures 3: Models performance: Cases A, B, C, and D (expected percentage correct),

with lower and upper limits of accuracy for SV (four panels on left) and TTB (four panels on right).

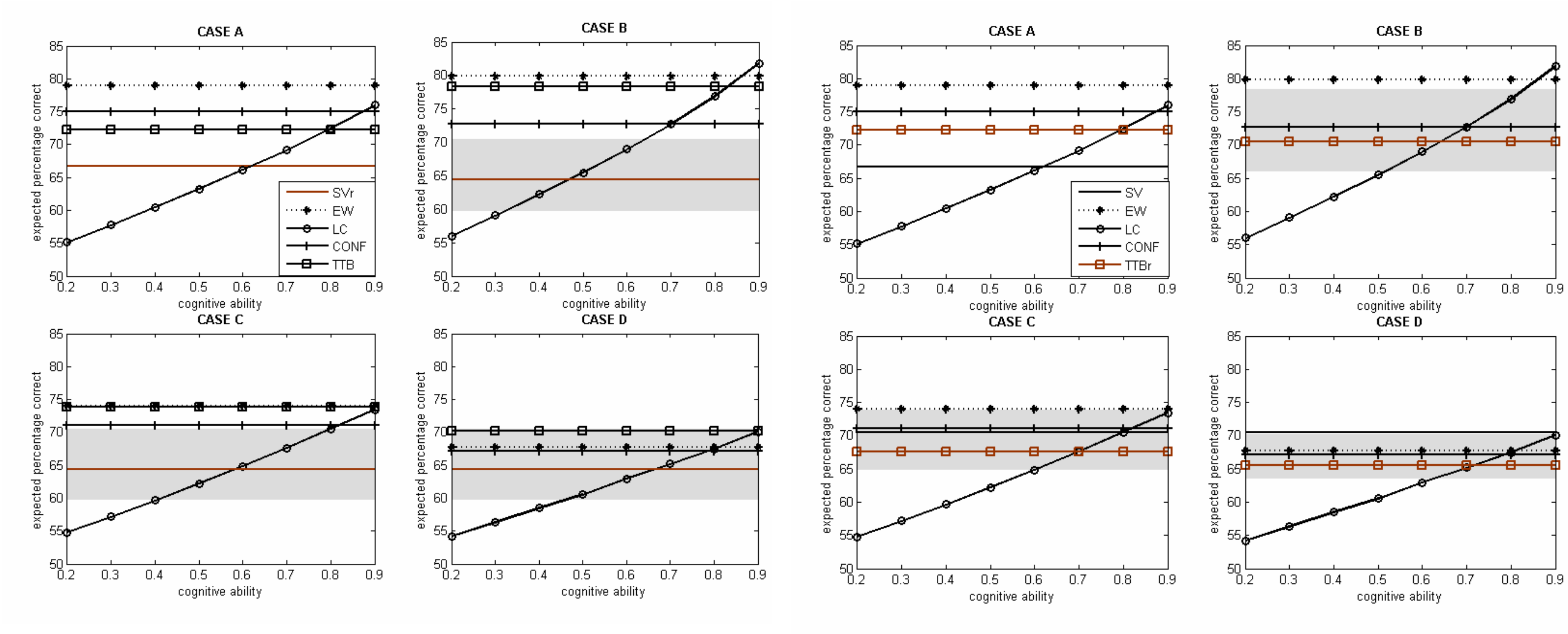
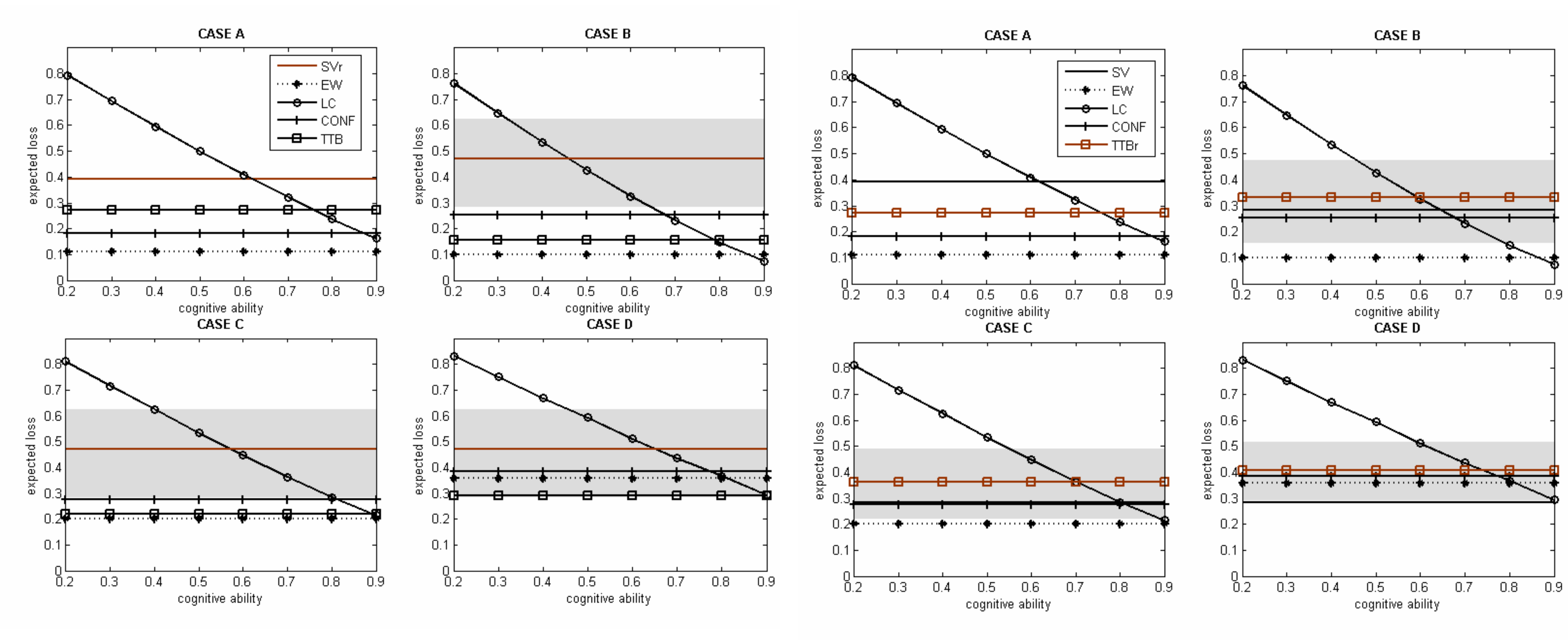


Figure 4: Models performance: Cases A, B, C, and D (expected loss for α=1.00),

with lower and upper limits of losses for SV (four panels on left) and TTB (four panels on right).



<u>Figure 5: Models performance: Cases A, B, C, and D (expected loss for α=0.30)</u>

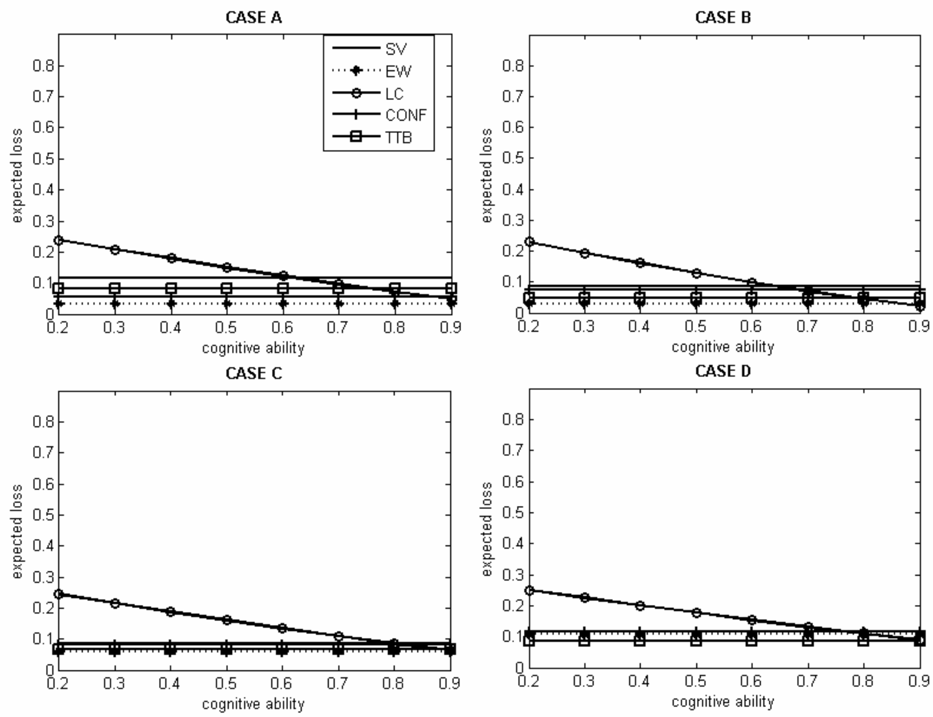


Figure 6: Models performance: Cases E, F, G, and H (expected percentage correct),

with lower and upper limits of accuracy for SV (four panels on left) and TTB (four panels on right).

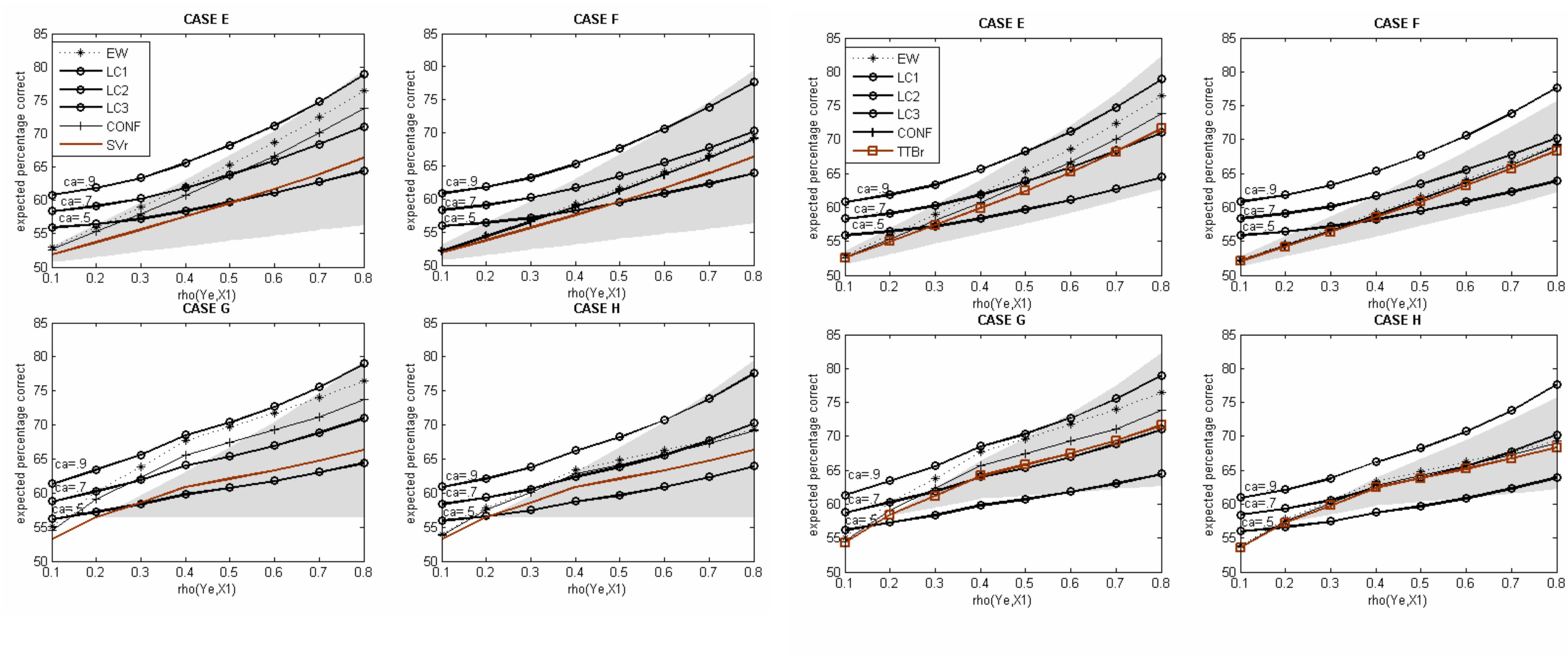
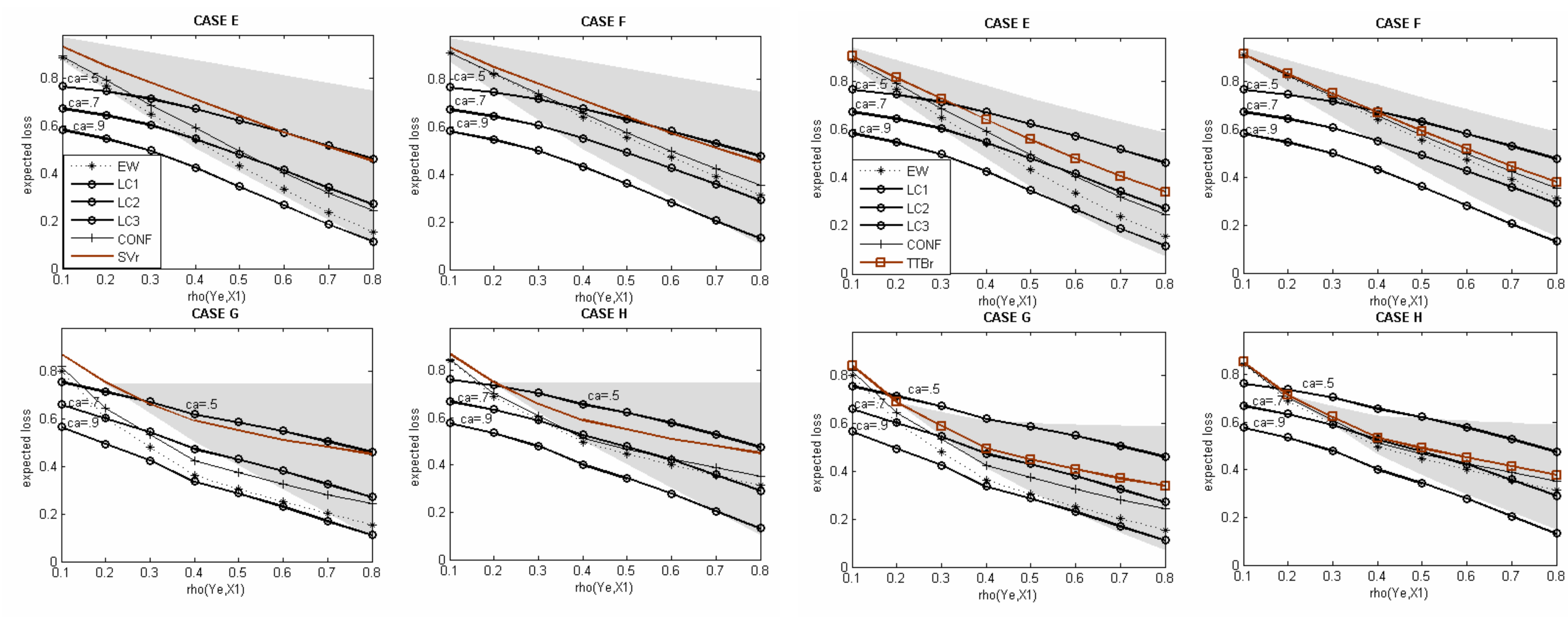


Figure 7: Models performance: Cases E, F, G, and H (expected loss for α=1.00)

with lower and upper limits of losses for SV (four panels on left) and TTB (four panels on right).



Appendix A – The expected accuracy of CONF and TTB.

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 probability of:

$$2\begin{bmatrix} P\{(Y_{ea} > Y_{eb}) \cap (X_{1a} > X_{1b}) \cap (X_{2a} > X_{2b})\} + \\ P\{(Y_{ea} > Y_{eb}) \cap (X_{1a} > X_{1b}) \cap (X_{2a} < X_{2b}) \cap (X_{3a} > X_{3b})\} + \\ P\{(Y_{ea} < Y_{eb}) \cap (X_{1a} > X_{1b}) \cap (X_{2a} < X_{2b}) \cap (X_{3a} < X_{3b})\} \end{bmatrix} =$$

$$2\left[\int_0^\infty\int_0^\infty\int_0^\infty f_1(d)\ dd + \int_0^\infty\int_0^\infty\int_{-\infty}^0\int_0^\infty f_2(d)\ dd + \int_{-\infty}^0\int_0^\infty\int_{-\infty}^0\int_{-\infty}^0 f_2(d)\ dd\right] \qquad \text{(A1)}$$

where both $f_1(d) = f_1(d_1, d_2, d_3)$ and $f_2(d) = f_2(d_1, d_2, d_3, d_4)$ are defined by $\frac{\left|M_f^{-1}\right|^{1/2}}{2\pi} e^{-\frac{1}{2}d' M_f^{-1} d}$, the variance/ covariance matrix specific to each being:

$$M_{f_1} = \begin{pmatrix} 2 & 2\rho_{Y_e X_1} & 2\rho_{Y_e X_2} \\ 2\rho_{Y_e X_1} & 2 & 2\rho_{X_1 X_2} \\ 2\rho_{Y_e X_2} & 2\rho_{X_1 X_2} & 2 \end{pmatrix} \text{ and } M_{f_2} = \begin{pmatrix} 2 & 2\rho_{Y_e X_1} & 2\rho_{Y_e X_2} & 2\rho_{Y_e X_3} \\ 2\rho_{Y_e X_1} & 2 & 2\rho_{X_1 X_2} & 2\rho_{X_1 X_3} \\ 2\rho_{Y_e X_2} & 2\rho_{X_1 X_2} & 2 & 2\rho_{X_2 X_3} \\ 2\rho_{Y_e X_3} & 2\rho_{X_1 X_3} & 2\rho_{X_2 X_3} & 2 \end{pmatrix}.$$

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

$$2\begin{bmatrix} P\{(Y_{ea} > Y_{eb}) \cap (X_{1a} - X_{1b} > t)\} + \\ P\{(Y_{ea} > Y_{eb}) \cap (|X_{1a} - X_{1b}| < t) \cap (X_{2a} - X_{2b} \geq t)\} + \\ P\{(Y_{ea} > Y_{eb}) \cap (|X_{1a} - X_{1b}| < t) \cap (|X_{2a} - X_{2b}| < t) \cap (X_{3a} - X_{3b} \geq t)\} \end{bmatrix} +$$

$$P\{(Y_{ea} > Y_{eb}) \cap (|X_{1a} - X_{1b}| < t) \cap (|X_{2a} - X_{2b}| < t) \cap (|X_{3a} - X_{3b}| < t)\} =$$

$$2\left[\int_0^\infty \int_t^\infty f_3(d)\ dd + \int_0^\infty \int_{-t}^t \int_t^\infty f_1(d)\ dd + \int_0^\infty \int_{-t}^t \int_{-t}^t \int_t^\infty f_2(d)\ dd\right] + \int_0^\infty \int_{-t}^t \int_{-t}^t \int_{-t}^t f_2(d)\ dd \quad \text{(A2)}$$

where both $f_1(d) = f_1(d_1, d_2, d_3)$ and $f_2(d) = f_2(d_1, d_2, d_3, d_4)$ are the same as in CONF, and $f_3(d) = f_3(d_1, d_2)$ is found 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 CONF and TTB

The expected loss of **CONF** is:

$$2L\begin{bmatrix} P\{(Y_{ea}<Y_{eb})\cap(X_{1a}>X_{1b})\cap(X_{2a}>X_{2b})\}+ \\ P\{(Y_{ea}<Y_{eb})\cap(X_{1a}>X_{1b})\cap(X_{2a}<X_{2b})\cap(X_{3a}>X_{3b})\}+ \\ P\{(Y_{ea}>Y_{eb})\cap(X_{1a}>X_{1b})\cap(X_{2a}<X_{2b})\cap(X_{3a}<X_{3b})\} \end{bmatrix}=$$

$$2\alpha\left[\int_{-\infty}^{0}\int_{0}^{\infty}\int_{0}^{\infty} d_1^2 f_1(d)\ dd+\int_{-\infty}^{0}\int_{0}^{\infty}\int_{-\infty}^{0}\int_{0}^{\infty} d_1^2 f_2(d)\ dd+\int_{0}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{0}\int_{-\infty}^{0} d_1^2 f_2(d)\ dd\right] \quad \text{(B1)}$$

with $f_1(d)$ and $f_2(d)$ are as defined in Appendix A.

The expected loss of **TTB** is:

$$2L\begin{bmatrix} P\{(Y_{ea}<Y_{eb})\cap(X_{1a}-X_{1b}>t)\}+ \\ P\{(Y_{ea}<Y_{eb})\cap(|X_{1a}-X_{1b}|<t)\cap(X_{2a}-X_{2b}\geq t)\}+ \\ P\{(Y_{ea}<Y_{eb})\cap(|X_{1a}-X_{1b}|<t)\cap(|X_{2a}-X_{2b}|<t)\cap(X_{3a}-X_{3b}\geq t)\} \end{bmatrix}+$$
$$LP\{(Y_{ea}<Y_{eb})\cap(|X_{1a}-X_{1b}|<t)\cap(|X_{2a}-X_{2b}|<t)\cap(|X_{3a}-X_{3b}|<t)\}=$$

$$\alpha\left(2\left[\int_{-\infty}^{0}\int_{t}^{\infty} d_1^2 f_3(d)\ dd+\int_{-\infty}^{0}\int_{-t}^{t}\int_{t}^{\infty} d_1^2 f_1(d)\ dd+\int_{-\infty}^{0}\int_{-t}^{t}\int_{-t}^{t}\int_{t}^{\infty} d_1^2 f_2(d)\ dd\right]+\int_{-\infty}^{0}\int_{-t}^{t}\int_{-t}^{t}\int_{-t}^{t} d_1^2 f_2(d)\ dd\right)$$
(B2)

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

Appendix C – Selected 3-cue studies.

No.	Study	Task	Number of conditions/tasks	Total number of participants	Stimuli per participant	R_e across conditions (range)	across conditions (range) r_a	GRs
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 et al. (1999)	Artificial prediction task	5	220	120	0.57-0.98	0.37-0.78	0.67-0.82
Compensatory environments								
4	Chasseigne et al. (1977) - Experiment 1	Artificial prediction task	6	96	26	0.96	0.34-0.70	0.35-0.73
5	Kessler & Ashton (1981)	Prediction of corporate bond ratings	4	69	34	0.74	0.52-0.64	0.71-0.88
6a*	Steinmann (1974)	Artificial prediction task	9	11	300	0.63-0.78	0.45-0.57	0.68-0.84
Non-compensatory environments								
2b	Brehmer & Hagafors (1986)	Artificial prediction task	2	20	15	0.77-1.00	0.74-0.78	0.71-0.75
7	Deane et al. (1972) - Experiment 2	Artificial prediction task	2	40	20	0.94	0.59-0.84	0.65-0.89
8	Hammond et al. (1973)	Artificial prediction task	3	30	20	0.92	0.05-0.78	0.14-0.83
9	Hoffman et al. (1981)	Artificial prediction task	9	182	25	0.94	0.09-0.71	0.15-0.78
6b*	Steinmann (1974)	Artificial prediction task	6	11	100	0.63-0.74	0.44-0.65	0.70-0.85
10	Youmans & Stone (2005)	Prediction of income levels	4	117	50	0.44	0.35-0.42	0.88-0.97
		Total	54	944				

Notes:

1. All studies reported involved between-subject designs unless studies No. 6a & 6b (indicated by *).
2. Three studies -- No. 7, 8, and 9 -- were said to have identical parameters. However, there must have been some rounding differences because of marginally different values reported for Re.

Appendix D – Selected 2-cue studies.

No.	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	GRs
Equal weighting environments								
1	Jarnecke & Rudestam (1976)	Predict academic achievement	1	15	50	0.42	0.28	0.71
2	Lafon et al. (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 et al. (1969)	Judging the age of blood cells	1	16	64	0.99	0.73	0.73
Non-compensatory environments								
5	Armelius & Armelius (1974)	Artificial prediction task	3	63	25	0.99-1.00	0.32-0.96	0.32-0.95
6	Doherty et al. (1988)	Artificial prediction task						
		Experiment 2	3	45	25	0.79-1.00	0.70-0.73	0.74-0.92
		Experiment 6	2	30	50	0.87-1.00	0.53-0.66	0.58-0.73
7	Hammond & Summers (1965)	Artificial prediction task	3	30	20	0.71	0.49-0.85	0.48-0.59
8	Lee & Yates (1992)	Post-dicting student success	2	40	NA	0.38	0.24-0.29	0.51-0.59
9	Muchinsky & Dudycha (1975)	Artificial prediction task						
		Experiment 1	2	160	150	0.72	0.04-0.30	0.11-0.54
		Experiment 2	2	160	150	0.96	0.03-0.45	0.01-0.32
10	Steinmann & Doherty (1972)	Assessing subjective probabilities in a bookbag and poker chip task	1	22	192	0.95	0.67*	0.70*
11	York et al. (1987)	Artificial prediction task	3	45	25	0.86	0.53-0.64	0.62-0.74
		Total	33	1137				

Notes:

1. The number of participants in studies No. 3 and 8 are approximations since this information is not available.
2. In study No. 10, "human performance" was measured through medians (marked with *).