Available online at www.sciencedirect.com

SCIENCE DIRECT®

Organizational Behavior and Human Decision Processes 95 (2004) 1–19

ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES

www.elsevier.com/locate/obhdp

A new integrated model of noncompensatory and compensatory decision strategies☆

Terry Elrod,* Richard D. Johnson, and Joan White

3-23 School of Business, University of Alberta, Edmonton, Alta., Canada T6G 2R6

Received 18 August 2002

Abstract

We describe and test a model that captures conjunctive, disjunctive, and compensatory judgment and choice strategies, as well as selected hybrid combinations of these. This model: (a) can be estimated solely from nonexperimental outcome data, (b) remains true to the conceptualization of noncompensatory heuristics as cognitively less demanding for decision makers, (c) is truly noncompensatory and not just approximately, (d) tests for a "pervasive" influence of cutoffs, (e) allows for the possibility that decision makers use different strategies across attributes, and (f) provides a more plausible account of behavior than competing models. We show empirically that decision makers may sometimes devalue objects for almost failing a conjunctive criterion or value objects more favorably for almost passing a disjunctive criterion—what we term a pervasive influence of a cutoff. The superiority of the proposed model relative to two other state-of-the-art models is demonstrated using both actual admit/reject decisions of an MBA admissions office as well as 10 simulations of various decision tasks.

Keywords: Noncompensatory evaluation; Compensatory evaluation; Choice behavior; Additive models; Multiattribute decision making

Introduction

It is a well-known irony in judgment and decision making (J/DM) research that decision strategies that are cognitively more demanding can be simpler to model than less demanding strategies. For example, the linear compensatory model, which implies a moderate amount of information processing, is readily estimated by linear regression. On the other hand, the conjunctive decision rule, which entails rejection of any object that fails to meet a minimum criterion on an attribute, is simpler for decision makers to implement and yet it is harder for researchers to model in estimable form.

A valid mathematical model of decision making is important because it provides a precise specification of theory. It is also of practical importance because it allows researchers to infer the unobserved decision strategy from the observed decision outcomes, eliminating reliance on self-reports (e.g., Johnson, 1987; Swait, 2001), protocol data (e.g., Billings & Marcus, 1983; Klein & Bither, 1987; Payne, Bettman, & Johnson, 1988), or multiple observations of intermediate steps (Levin & Jasper, 1995). This is of value to practitioners, as well as academic researchers, because process data are often unavailable. Process data can also be unreliable. Self-reports are suspect because subjects may be unaware of their own decision strategies or unable to report them accurately (e.g., Nisbett & Wilson, 1977). Methods for collecting protocol data, such as information display boards, can interfere with the decision process they measure in several ways (e.g., Billings & Marcus, 1983; Ford, Schmitt, Schechtman, Hults, & Doherty, 1989). For example, they may (a) induce subjects to process information more carefully when their decision process is being monitored, (b) direct subjects' attention to attributes that would be ignored in a naturalistic setting (Brucks, 1985), and/or (c) cause information overload.

☆ The first author acknowledges support from a Winspear Endowed Senior Faculty Fellowship, School of Business, University of Alberta.

* Corresponding author. Fax: +1-780-492-3325.
E-mail address: terry.elrod@ualberta.ca (T. Elrod).

0749-5978/$ - see front matter
doi:10.1016/j.obhdp.2004.06.002

Classic utility theory (Debreu, 1954; von Neumann & Morgenstern, 1947) offers an idealized representation of decision outcomes that is readily extended to preference for multiattribute objects. The utility of the object is modeled as a function of the valuations (utilities) of its component characteristics (Lancaster, 1966). Mathematical tractability in economic analysis is obtained by assuming that the utilities of objects are compensatory; i.e., that the valuation of every attribute affects the utility of the object, irrespective of the valuations of the object's other attributes.

Theoretical critiques of classic utility theory date back at least to Mosteller and Nogee (1951), Simon (1955), and Luce (1956). These researchers embarked on a search for (a) equally general models that (b) provide more accurate accounts of human behavior while (c) presuming less extreme complexity of evaluation. Attaining all three goals simultaneously is no easy task, and subsequent research has emphasized some objectives more than others.

Some researchers reduced the compensatory model to its simplest possible form. The linear compensatory (or additive) model soon came to dominate behavioral research in multiattribute evaluation and choice behavior (Edwards & Tversky, 1967). Mosteller and Nogee (1951) noted that observed choices at the individual level frequently depart from deterministic utility theory, leading to the addition of a random error term to utilities and hence to *random utility models* (Block & Marschak, 1960; McFadden, 1974, 1987). McFadden's (1974) *conditional multinomial logit model*—a particular form of linear compensatory random utility model—likely remains the most often applied model of decision making to this day (Borooah, 2002). It is easy to estimate, and random utility models have proven hard to beat in terms of predictive accuracy, particularly in natural settings.

Other researchers have proposed noncompensatory rules thought to be even simpler for decision makers to implement than the linear compensatory rule. A decision rule is noncompensatory if a decision, determined by some attributes of an object, *cannot* be reversed by other attributes of the object (Dillon, 1998; Schoemaker, 1980). Noncompensatory models of enduring interest include the conjunctive and disjunctive models (Coombs & Kao, 1955; Dawes, 1964), the lexicographic semiorder (Tversky, 1969), elimination by aspects (Tversky, 1972) and fast and frugal heuristics (Gigerenzer, 2000). However, these models also have their shortcomings. (a) They preclude compensatory behavior, even though there is good reason to suppose that decision makers may use both compensatory and noncompensatory rules on different occasions or even at different stages in the same task (Beach, 1993; Payne, 1976; Payne, Bettman, & Luce, 1998). (b) Statistical procedures for fitting these models to nonexperimental outcomes are either not provided or else are problematic. For example, the number of parameters estimated by the models of Tversky (1969, 1972) and Gigerenzer (2000) is not well-defined, which precludes the statistical testing of hypotheses and hampers model assessment. (c) The domain of applicability of these models can be restrictive. For example, both fast and frugal heuristics and elimination by aspects were developed for binary attributes. (d) While the investigators of these models have regularly constructed task conditions for which noncompensatory models provide superior predictions, it is still unclear to what extent these conditions prevail in real world settings.

Some J/DM researchers have proposed models meant to represent noncompensatory decision strategies that can be readily estimated from nonexperimental outcomes (e.g., Abe, 1999; Brannick, 1986; Brannick & Brannick, 1989; Einhorn, 1970, 1971; Ganzach & Czaczkes, 1995; Goldberg, 1971; Mela & Lehmann, 1995; Slovic & Lichtenstein, 1971). While of value, these models all possess at least one of several important weaknesses: (a) they are inherently nonlinear compensatory models that can only approximate noncompensatory processes, (b) they assume that the same decision strategy is applied to all attributes, and/or (c) they implicitly represent the noncompensatory process as a complex consideration of either attribute interactions or of the means and variances of attribute values. Since these models suggest significantly more, rather than less, processing than is required by linear compensatory models, they are unrepresentative of the simple cognitive processes they seek to represent.

One extant type of model that can represent both linear compensatory and noncompensatory rules exactly is the *connectionist network*, or neural net (Ripley, 1996). Researchers in behavioral decision theory have been quick to take advantage of the flexibility of connectionist networks to model decision making—examples are to be found in Grossberg (1980), Grossberg and Gutowski (1987), Leven and Levine (1996), and in *decision field theory* (Busemeyer & Diederich, 2002; Busemeyer & Townsend, 1993; Roe, Busemeyer, & Townsend, 2001). However connectionist networks either make no probabilistic predictions for observations or else they make probabilistic predictions confined to the open interval (0,1) for all observations.[1] In the former case, they contain no error theory and so precise statistical assessment and hypothesis testing is not possible. In the latter case, they are inconsistent with behavioral descriptions of noncompensatory decision making because all object attributes affect evaluation and no object is ever rejected or accepted with certainty. Nevertheless these models can provide for a more ade-

[1] This, according to whether they model outputs using a threshold or sigmoid function, respectively (cf. Cherkassky & Mulier, 1998). The *open* interval (0, 1), excludes its endpoints of zero and one, whereas a *closed* interval, denoted [0, 1], includes them.

quate account of observed behavior in some settings than the random utility model. We will return to these models in the Discussion (Incorporation into decision field theory).

Advantages of the proposed model

In this paper, we propose and test a model of decision making that integrates variations of a compensatory and two noncompensatory (i.e., conjunctive and disjunctive) decision strategies. It is capable of providing probabilistic predictions for objects anywhere on the closed interval [0,1].

Our model offers a number of important advantages for decision researchers. (a) It allows for the identification of compensatory and noncompensatory decision strategies from experimental or nonexperimental choices, eliminating the need for protocol data or self-reports. (b) It remains true to the conceptualization of noncompensatory heuristics as being cognitively less demanding than compensatory decision making. (c) It allows for the possibility that decision makers may use a compensatory strategy for some attributes, conjunctive or disjunctive for others, and a combination of compensatory and noncompensatory for still other attributes, all within a single decision. (d) It tests for a *pervasive* influence of cutoffs, where objects that do not quite invoke a conjunctive or disjunctive rule receive evaluations that are nonetheless influenced by that rule. That is, objects that nearly fail a conjunctive criterion may be devalued, and those that nearly satisfy a disjunctive criterion may be valued more highly, relative to objects that do not nearly invoke the rule. Our empirical tests show that the decision strategies identified by our model are more consistent with theory and more plausible than the accounts of behavior that are implied by state-of-the-art alternative models.

The remainder of the Introduction expands on each of these points as we present the derivation of our model. We begin by characterizing in precise terms the linear compensatory and two common noncompensatory decision rules, and then show how these may be integrated into a single model. We present and assess a random utility implementation of this model to decisions about individual objects. We show in the Discussion how our core contribution—the additive GNH model—may be employed in more flexible choice models and more general choice tasks.

The linear compensatory model

Utility-based decision models assume that decisions are based upon evaluations, i.e., that decision makers seek options that offer high perceived utility. We will denote the utility of the ith object by U_i, and the object's values on the Q attributes that affect assessment by $X_{i1}, \ldots, X_{iQ}$. We assume that all attributes are coded to be positively valued; i.e., that the expected utility of an object is monotonically nondecreasing in every attribute. Negatively valued attributes can always be made positively valued by multiplying their data values by minus one. Furthermore, utility will often be latent rather than observed.

The most common representation of the decision maker's process of object evaluation is the linear compensatory model (Brunswik, 1940):

$$U_i = \alpha + \sum_{q=1}^{Q} \beta_q X_{iq} + \varepsilon_i, \tag{1}$$

where α is an intercept, ε_i is a random error term, and β_q is the influence of the qth attribute on the utility assessment.

The ability of the linear model to predict actual decisions reasonably well in many (but not all) settings is well documented (Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Slovic & Lichtenstein, 1971; Yntema & Torgerson, 1961). Its robust performance, attributable to a combination of completeness in its representation of attributes and simplicity in its combination rule, explains why it lies at the core of J/DM models to this day (e.g., Busemeyer & Diederich, 2002; Busemeyer & Townsend, 1993).

The linear model has three key features well-known to J/DM researchers. It is *additive in attributes*, which means that the evaluation of the object is obtained by simply summing the assessments of each of the attributes considered individually. It is *compensatory*, which implies that an object's assessment on any attribute may be offset by its assessment on one or more other attributes. And finally, it is *linear*, which means that all attributes are assessed in a linear manner for all objects.

Common variations on this rule relax the additivity and/or linearity assumptions. Models that are additive in attributes can be expressed as

$$U_i = \alpha + \sum_{q=1}^{Q} V_{iq} + \varepsilon_i \equiv V_i^* + \varepsilon_i, \tag{2}$$

where $V_{iq} \equiv f_q(X_{iq})$ represents the possibly nonlinear *value function* for the qth attribute of object i and ε_i is a random error term. It is sometimes useful to separate U_i into its deterministic (V_i^*) and random (ε_i) components. The linear compensatory model given by (1) is a special case of (2) with $V_{iq} = \beta_q X_{iq}$.

Noncompensatory rules

Noncompensatory decision strategies differ from linear compensatory rules in that a decision may be determined by an object's score on a single attribute, irrespective of its score on other attributes (e.g., Coombs & Kao, 1955; Dawes, 1964; Pras & Summers, 1975; Schoemaker, 1980). For example, a house that is unaffordable is rejected regardless of its features or location. This is an example of a conjunctive decision rule, in which an object is rejected because it fails to meet a minimum level of desirability for at least one of its attributes. Alternatively, a disjunctive rule results in acceptance of an object that surpasses a very high standard on at least one attribute, irrespective of its values on the other attributes. For example, one might accept the first job offer in a specific location without regard for salary or working conditions.

Letting $Y_i = 2$ indicate an "acceptable" object and $Y_i = 1$ an "unacceptable" one, then the conjunctive rule may be represented as

$$Y_i = 1 \text{ if and only if } X_{iq} < \delta_q \text{ for any } q = 1, \ldots, Q \quad (3)$$

and the disjunctive rule as

$$Y_i = 2 \text{ if and only if } X_{iq} > \delta_q \text{ for any } q = 1, \ldots, Q. \quad (4)$$

With both conjunctive and disjunctive rules, the decision maker need only compare the object's value on an attribute to the cutoff criterion for that attribute (δ_q).

Noncompensatory decision making can be significantly more efficient than full-information processing (Payne, Bettman, & Johnson, 1993). In addition, it may be used to reduce the number of objects to be evaluated more carefully (Beach, 1993; Payne, 1976; Payne et al., 1998). Processing efficiency is important to decision makers because the number of objects and attributes that can be considered is severely constrained by human working memory capacity and computational ability (e.g., Miller, 1956, Shugan, 1980).

The pervasive influence of noncompensatory cutoffs

The cutoff in conjunctive and disjunctive decision strategies has traditionally been characterized as a single attribute value; for example, a firm that is seeking to fill a managerial position rejects all candidates with less than 5 years of work experience, regardless of their education or other qualifications. It is assumed that the cutoff has no effect on the evaluation of candidates with more than 5 years of experience. In contrast, we allow for the possibility that the cutoff may have a *pervasive* influence, affecting the evaluation of objects that come close to invoking the noncompensatory rule. For example, candidates that nearly fail a conjunctive cutoff (e.g., candidates with five and one-half years of experience) may be evaluated less favorably than candidates that pass the cutoff by a comfortable margin. Similarly, candidates that are nearly, but not quite, accepted according to a disjunctive rule may be evaluated more favorably, and thus be more likely chosen, than candidates that do not come close to the disjunctive cutoff.

Sometimes a noncompensatory rule may be used to reduce the number of objects to be evaluated in a linear compensatory manner. In such cases, pervasive influence would imply that objects close to a conjunctive (/disjunctive) cutoff would receive more unfavorable (/favorable) evaluations than would result from linear compensatory evaluation.

Other authors have proposed that decision makers sometimes violate *self-reported* cutoffs (Green, Krieger, & Bansal, 1988; Huber & Klein, 1991; Johnson, 1987; Swait, 2001). Both Johnson and Swait describe models that penalize, but do not eliminate, objects that fail to meet a stated conjunctive cutoff. In fact, both of these models are compensatory approximations to noncompensatory behavior. Swait uses self-reported cutoffs that, unless adhered to for every choice by every respondent, merely serve to locate points of nonlinearity in an attribute value function that is compensatory throughout its range. Johnson handles stated cutoffs in a similar, compensatory, manner.

On the other hand, the model proposed here relies on observed choices and makes no use of self-reported cutoffs. Thus any pervasive effects of cutoffs are genuine and not an artifact of a self-report procedure. Furthermore, the model treats noncompensatory cutoffs as truly noncompensatory—objects that invoke a conjunctive (/disjunctive) rule are still rejected (/accepted) with certainty, regardless of the other attributes of these objects.

A new integrated model of decision strategies

Our objective is a model that (a) remains true to the well-established belief that noncompensatory heuristics are used by decision makers to simplify the decision making task while (b) providing a more plausible account of behavior than closely competing models. We accomplish this by adapting the additive model to include noncompensatory evaluation.

Noncompensatory additive models

It might appear that additive multiattribute utility functions (cf. (2)) necessarily imply compensatory decision making, but this is not so. Consider the case of a conjunctive rule, which implies that an object is rejected if it fails to meet minimum cutoff(s) on any attribute(s) regardless of its values on other attributes. This rule, given by (3), can be represented exactly using the additive model of (2) by setting α equal to a very large positive constant (i.e., $\alpha \gg \sigma$, where σ is the standard deviation of the error term), letting

$$V_{iq} = \begin{cases} 0, & X_{iq} \geqslant \delta_q \\ -\infty, & X_{iq} < \delta_q \end{cases}, \quad q = 1, \ldots, Q, \quad (5)$$

and using the binary decision rule

$$Y_i = \begin{cases} 2, & U_i \geqslant 0, \\ 1, & U_i < 0. \end{cases} \quad (6)$$

An object will be selected with certainty[2] if it passes all the criteria of (5); otherwise it will be rejected with certainty.

Similarly, the disjunctive rule results from using (2) with α set to a very large negative constant (i.e., $\alpha \ll -\sigma$) together with the binary decision rule given by (6) and the attribute value function

$$V_{iq} = \begin{cases} +\infty, & X_{iq} > \delta_q \\ 0, & X_{iq} \leqslant \delta_q \end{cases}, \quad q = 1, \ldots, Q. \tag{7}$$

In this case, objects that exceed one or more disjunctive criteria are selected with certainty, and the rest are rejected with certainty.[2]

Hybrid valuation

We have noted that conjunctive rules in particular may be an effective means for reducing the number of objects submitted to more careful (e.g., linear compensatory) evaluation (cf. Beach, 1993; Gensch, 1987; Johnson, 1987; Levin & Jasper, 1995; Payne, 1976; Payne et al., 1998; Pras & Summers, 1975; Roberts, 1988). By simply adding the term $\beta_q X_{iq}$ to the conjunctive value function of (5), the proposed model is able to uncover evidence of linear compensatory processing of objects that pass conjunctive criteria in this manner. Objects that fail the conjunctive criteria are still rejected with certainty but objects that are not eliminated receive valuations on that attribute according to the linear compensatory model. The same addition of the term $\beta_q X_{iq}$ to the disjunctive value function of (7) results in linear compensatory valuation of objects that are not accepted outright by the disjunctive rule. Since the order of processing cannot be determined when only decision outcomes are observed, the model can identify when both conjunctive/disjunctive and compensatory rules are used, but it cannot determine whether the rules are applied sequentially or simultaneously.

Illustrations of decision strategies represented by the proposed model

Examples of conjunctive, linear, and conjunctive–linear rules are illustrated in Fig. 1 along with two other rules that represent a pervasive effect of a cutoff. The choices of scale for both attribute value X and its valuation V are arbitrary. The linear compensatory rule is illustrated by the line of points labeled "+." The conjunctive rule, referred to as "crisp conjunctive" in the figure, is represented by the points labeled "○." According to this rule, objects with $X \geqslant 2$ (in this example) are valued at 0 on this attribute, whereas all other objects are valued at minus infinity. The hybrid conjunctive–linear compensatory rule, denoted "crisp conjunctive plus linear," is simply the sum of the crisp conjunctive and linear curves and is represented by the thin solid line. In Fig. 2, the same symbols portray the analogous rules for the disjunctive case.

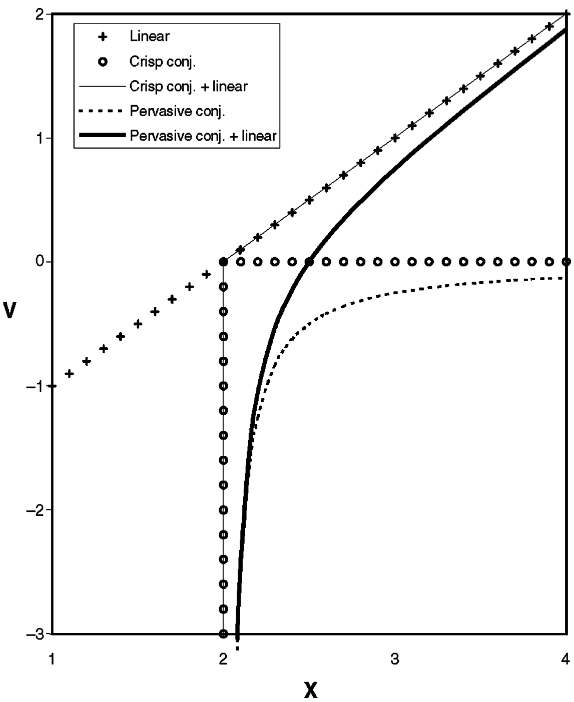

Fig. 1. Illustrations of crisp/pervasive conjunctive and/or linear value functions.

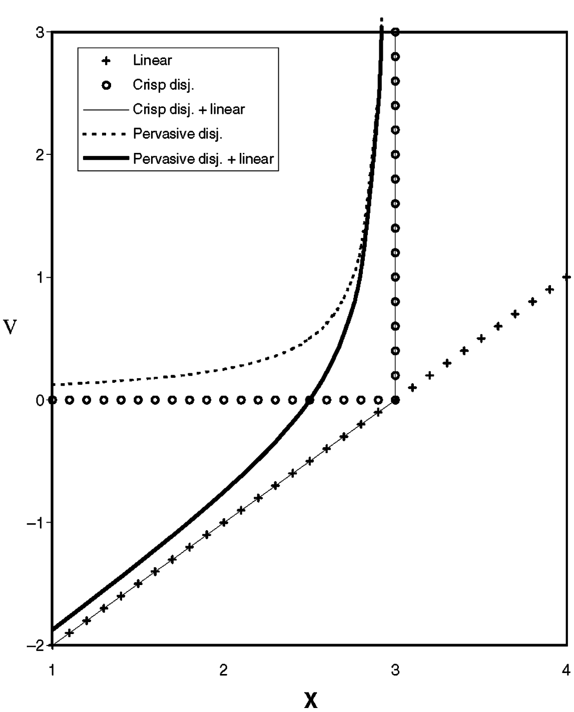

Fig. 2. Illustrations of crisp/pervasive disjunctive and/or linear value functions.

[2] To machine precision.

The dashed line in Fig. 1 illustrates an implementation of a pervasive conjunctive rule that retains its noncompensatory characteristic. Objects that nearly fail the conjunctive criterion are valued more negatively than objects that easily satisfy it. The closer the object comes to failing, the more negative the pervasive effect. The dashed curve shown in Fig. 1 happens to be highly pervasive. An estimated "pervasiveness" parameter determines the extent to which the pervasive conjunctive rule departs from (i.e., "rounds") the crisp conjunctive rule.

The final rule illustrated in Fig. 1 is the pervasive conjunctive-plus-linear rule, shown as a bold solid line. Again, this rule is obtained by simply summing the pervasive-conjunctive and linear curves. The corresponding rules for the disjunctive case are illustrated in Fig. 2. Notice that, of the nine different curves shown in Figs. 1 and 2 (the linear rule is shown twice), only the linear compensatory rule is compensatory for all values of X.

The general nonrectangular hyperbola

The value function discussed and tested in this paper is sometimes referred to as the general nonrectangular hyperbola (GNH) (Ross, 1990, p. 154), which involves the estimation of up to three parameters for each numeric variable included in the analysis. The GNH generated all of the curves displayed in Figs. 1 and 2.

The GNH for the *conjunctive* case may be expressed as

$$V_{iq} = \begin{cases} \frac{-\gamma_q}{X_{iq}-\delta_q} + \beta_q X_{iq}, & X_{iq} \geqslant \delta_q \\ -\infty, & X_{iq} < \delta_q \end{cases} \tag{8}$$

and for the *disjunctive* case as

$$V_{iq} = \begin{cases} \frac{-\gamma_q}{X_{iq}-\delta_q} + \beta_q X_{iq}, & X_{iq} \leqslant \delta_q, \\ +\infty, & X_{iq} > \delta_q. \end{cases} \tag{9}$$

Fortunately, it is possible to combine the conjunctive and disjunctive cases into a single rule. This saves researchers the inconvenience of having to try all combinations of formulae for all of the continuous variables and also simplifies model comparison and testing. Obtaining a single rule is possible because both (8) and (9) approach the linear case, even when $\beta_q = 0$, as δ_q approaches minus infinity for the conjunctive case and as δ_q approaches plus infinity for the disjunctive case. Therefore, we estimate for both rules a parameter $-1 \leqslant \rho_q \leqslant +1$. Negative values of ρ_q imply the conjunctive case, positive values imply the disjunctive, and the value $\rho_q = 0$ implies the linear case $V_{iq} = \beta_q X_{iq}$. The *GNH value function*, which depends on ρ_q, is given by

$$\begin{aligned} &-1 \leqslant \rho_q < 0 \Rightarrow V_{iq} \text{ given by (8) with} \\ &\qquad \delta_q = 1/\tan(\rho_q \pi/2) + D_{\max,q} \\ &\rho_q = 0 \Rightarrow V_{iq} = \beta_q X_{iq} \qquad (10) \\ &0 \leqslant \rho_q < +1 \Rightarrow V_{iq} \text{ given by (9) with} \\ &\qquad \delta_q = 1/\tan(\rho_q \pi/2) + D_{\min,q} \end{aligned}$$

$D_{\max,q}$ is the maximum possible value for the conjunctive cutoff, which is the minimum value of X_q among accepted observations. Similarly, $D_{\min,q}$ is the minimum possible value for the disjunctive cutoff, which is the maximum value of X_q among rejected observations. Thus both $D_{\max}$ and $D_{\min}$ are observed for all attributes.

Because the GNH value function can represent several different strategies, the researcher need not know a priori what decision strategy underlies the evaluation of each attribute. However, the full generality of the three-parameter functional form may not be appropriate for every attribute. We specify in Table 1 special cases of the general function that involve the estimation of fewer parameters and represent each of the curves shown in Figs. 1 and 2. Tests of statistical significance or a model selection criterion may be used to determine the degree of generality required for each attribute. The *GNH model* explains object valuation (V_i^* of (2)) as the sum of an intercept (α, which may be zero) and the results of applying the GNH value function, or special cases of it (cf. (10) and Table 1), to each of the object's attributes.

Modeling the observed dependent variable

The additive model's ability to represent both compensatory and noncompensatory rules exactly is due to the distinction made between the latent evaluation of an object (U_i) and an observed decision made pertaining to that object (Y_i). We defer until the Discussion consideration of the case in which multiple objects are evaluated for each decision. Here, a decision is observed for each object. Two types of decisions are commonplace.

Binary evaluation/selection. The decision maker either accepts ($Y_i = 2$) or rejects ($Y_i = 1$) each object in turn. The relation between Y_i and U_i for this case is given by (6).

Ordinal evaluation. Often decision makers are asked to evaluate objects using a discrete and finite numeric rating scale, such as assigning to each object an integer between one and seven. The "decision" is choice of ordinal response that best reflects the evaluation. Such data are best modeled as ordinal. Letting S denote the number of points on the scale and using the coding $s \in \{1, \ldots, S\}$ for the ordinal responses, the model relating decision to evaluation is given by

$$Y_i = s \text{ if and only if } \phi_{s-1} \leqslant U_i < \phi_s,$$

where ϕ_0 is fixed at minus infinity, ϕ_1 at zero, ϕ_S at plus infinity, and the remaining values $\phi_2, \ldots, \phi_{S-1}$ (if any) are estimated subject to the requirement $\phi_{s-1} < \phi_s < \phi_{s+1}, s = 2, \ldots, S-1$.

Modeling stochastic decision making

So far we have shown how an additive model can exactly represent deterministic noncompensatory decision making. However, sample data are rarely deter-

Table 1
How the GNH value function represents each of nine decision rules

Rule	Implementation	Fixed parameters	Estimated parameters
Crisp conjunctive[a]	Eq. (8) with $\delta = D_{\max}$	$\gamma = 0,\ \beta = 0$	$\rho = -1$
Crisp conjunctive[a] plus linear	Eq. (8) with $\delta = D_{\max}$	$\gamma = 0$	$\rho = -1,\ \beta > 0$
Pervasive conjunctive	Eq. (8) with $\delta = 1/\tan(\rho\pi/2) + D_{\max}$	$\beta = 0$	$-1 < \rho < 0,\ \gamma > 0$
Pervasive conjunctive plus linear	Eq. (8) with $\delta = 1/\tan(\rho\pi/2) + D_{\max}$		$-1 < \rho < 0,\ \gamma > 0,\ \beta > 0$
Linear[b]	βX	$\rho = 0,\ \gamma = 0$	$\beta > 0$
Pervasive disjunctive plus linear	Eq. (9) with $\delta = 1/\tan(\rho\pi/2) + D_{\min}$		$0 < \rho < +1,\ \gamma > 0,\ \beta > 0$
Pervasive disjunctive	Eq. (9) with $\delta = 1/\tan(\rho\pi/2) + D_{\min}$	$\beta = 0$	$0 < \rho < +1,\ \gamma > 0$
Crisp disjunctive[c] plus linear	Eq. (9) with $\delta = D_{\min}$	$\gamma = 0$	$\rho = +1,\ \beta > 0$
Crisp disjunctive[c]	Eq. (9) with $\delta = D_{\min}$	$\gamma = 0,\ \beta = 0$	$\rho = +1$

[a] Crisp rules result from setting $\gamma = 0$. For crisp *conjunctive* rules, with or without a linear term, the maximum likelihood estimator of ρ is -1, which implies $\delta = D_{\max}$ in (8).

[b] The implementation $V = \beta X$ is obtained from (8) with $\delta = -\infty$ and $\gamma = 0$ and also from (9) with $\delta = +\infty$ and $\gamma = 0$. Both correspond to a value of $\rho = 0$.

[c] Crisp rules result from setting $\gamma = 0$. For crisp *disjunctive* rules, with or without a linear term, the maximum likelihood estimator of ρ is $+1$, which implies $\delta = D_{\min}$ in (9).

ministic. By specifying a distribution for the error term in (2), we can make probabilistic predictions for Y that lie anywhere in the closed interval [0,1].[3] Probabilistic formulation of a model is essential to model estimation and assessment. It allows us to determine those values for the unknowns in a model that maximize agreement between model and data (model estimation). It also provides us with a measure of inconsistency between model and data that permits model comparison and hypothesis testing (model assessment).

While assuming normality for ε leads to a conceptually appealing probabilistic model of decision making, we can employ a popular alternative that leads to simpler formulae for probabilistic predictions. In the case of binary or ordinal evaluation, the error term is often assumed to have the logistic distribution. In the ordinal case, the probability that $Y_i \leqslant s$ is equal to $1/[1 + \exp(V_i^* - \phi_s)]$, where V_i^* is the deterministic utility of the object (cf. (2)). Then the $\Pr(Y_i = s) = \Pr(Y_i \leqslant s) - \Pr(Y_i \leqslant s - 1)$. In the binary case this simplifies to

$$\Pr(Y_i = 2) = 1/[1 + \exp(-V_i^*)]. \tag{11}$$

Tests of the GNH model

In this section, we illustrate an application of the GNH model and assess it using both actual and simulated data. Both types of data support the superiority of the GNH model relative to both the linear compensatory model and two other additive models. "Superiority" is determined using both statistical and theoretical criteria. For the simulated data sets we also compare each estimated model to the known decision strategy used to generate the data.

Extant additive models

We compare our model with two nonlinear additive models that we believe come closest to representing hybrid compensatory–noncompensatory evaluation processes and that are widely employed in other contexts. These two models share with the GNH model the following advantages: (a) they can be estimated from decision outcomes without supplemental data on the decision process, (b) decision makers may apply either the same rule or different rules to the different attributes, (c) they represent conjunctive, linear compensatory, and disjunctive rules using a single formula, (d) they include the linear compensatory model as a special case, and (e) they avoid representing noncompensatory heuristics as complex manipulations of attribute means and variances or of attribute interactions.

However, the GNH model improves upon the two extant additive models in five respects. (a) It is derived from behavioral decision theory. (b) It represents the noncompensatory conjunctive and disjunctive decision rules [(3) and (4)] exactly, rather than approximately, providing estimates of the cutoff values they imply. (c) It allows for linear compensatory evaluation in conjunction with noncompensatory treatment of the same attribute as, for example, when a conjunctive rule is used to screen out objects and a compensatory model is applied to the resulting consideration set. (d) It allows the researcher to impose monotonicity of attribute evaluation on an attribute-by-attribute basis. (e) It tests for the pervasive influence of cutoffs.

[3] This is the random utility approach noted in the Introduction. We consider generalizations of this approach in the Discussion.

Polynomials

The ability of nonlinear compensatory additive models to mimic noncompensatory heuristics suggests that *polynomials* might be used to represent nonlinear evaluations of attributes. The polynomial value function for an attribute can be represented as

$$V_{iq} = \sum_{m=1}^{M_q} \beta_{qm}(X_{iq})^m, \quad q = 1, \ldots, Q, \tag{12}$$

where M_q is the number of polynomial terms used for the qth attribute.

While polynomials are nonlinear in attributes, they are linear in parameters, which means that estimation of polynomials is available in many different data analysis programs. The quadratic, corresponding to $M_q = 2$, may be the most useful polynomial because it often suffices to capture the most commonly observed nonlinearities. However, it implies that evaluations must be nonmonotonic, although the nonmonotonicity may occur outside the observed range of an attribute. Also, the degree of curvature that can be represented by a quadratic for monotonic attributes is limited, and monotonicity cannot be guaranteed. The cubic is better able to model differing degrees of curvature over the range of the data, but increased flexibility and imprecision of the estimates implies that nonmonotonicity is more likely to be inferred simply by chance. The cubic is also more sensitive to *influential* observations. That is, the shape of the curve at any point $X_q = x_0$ is determined by all the data, often primarily by data values far from x_0. These problems increase as more terms are used, with the result that polynomials of four or more terms are seldom estimated.[4]

Smoothing splines

Another means for fitting nonlinear compensatory models is the smoothing spline (Abe, 1999; Hastie & Tibshirani, 1990). *Ordinary* (not smoothing) splines (Schoenberg, 1946) are piecewise polynomial functions that use regularity conditions to reduce the number of parameters that must be estimated and increase the smoothness of the estimated function. The most common choice of ordinary spline is the natural cubic spline, which consists of cubic polynomials which must have equal heights and slopes at the "knots" at which they join. Splines are easy to estimate if the locations of the knots, which are specified values of X_q, are prespecified. However, the performance of the spline approximation depends upon how well the knots are chosen.

Smoothing splines (Hastie & Tibshirani, 1990) solve the problem of specifying the locations of knots for ordinary splines by making every observed value of X_q a knot. Such a large number of knots would allow an ordinary spline to fit the observed data perfectly, but would require the estimation of too many parameters for each attribute. A smoothing spline is fit by maximizing fit to the data while minimizing the curve's departure from a straight line. Increased smoothness is attained by increasing the penalty for curvature. Very smooth splines fit the data less well but are penalized less because they contain little curvature.

It is possible to obtain a good approximation of the degrees of freedom used up by any smoothing spline and, as implemented in S-Plus's `gam` function (Insightful Corp., 2001), the user may even specify how many degrees of freedom are to be used to estimate a smoothing spline relationship between the response and any independent variable. Thus, it is possible to use a model selection criterion or hypothesis tests to determine the degree of nonlinearity required to model the effect of each attribute. In addition, smoothing splines have the very desirable property that they are robust to influential observations—the shape of the curve at any point $X_q = x_0$ is always determined primarily by values of X_q close to x_0.

While the smoothing spline is an excellent tool for exploratory research, it must be regarded as a general purpose curve fitting procedure subject to two particular shortcomings as a tool for understanding evaluation and decision making. First, because the model's definition of smoothness is based on a straight line, it will never extrapolate curvature. For example, it regards a quadratic curve as less smooth than a quadratic curve that changes abruptly to a straight line. Second, it can provide a predicted value V_q for any value X_q, but it cannot provide an algebraic characterization of the attribute value function $V_q = f_q(X_q)$.

The statistical assessments described below corroborate the superiority of the GNH model to both the polynomial and the smoothing spline. Furthermore, the value functions recovered by the GNH model are more plausible and more consistent with behavioral theory.

A statistical measure of model performance

Our statistical measure of model performance is a model selection criterion (*MSC*) of the form:

$$MSC(\kappa) = -2L + \kappa P, \tag{13}$$

where L is the maximized loglikelihood of the estimated model, P is the number of parameters estimated, and κ, which must be positive, is the penalty incurred by a model for each parameter estimated. Better models have

[4] While the variables created by (12) can and should be transformed to orthogonality within attribute, correlation among attributes can become considerable, and often the data are insufficient to allow reliable estimation of four or more parameters per attribute (Ratkowsky, 1990). Furthermore, orthogonality is a least-squares property that eliminates dependence among estimates only when the response is amenable to analysis by ordinary regression.

lower MSC scores. Eq. (13) may be used to compare any number of models estimated by maximum likelihood, whether these models are nested or not.

Models that estimate more parameters tend to fit the data better, resulting in a smaller value for $-2L$, but they are also more heavily penalized because of the greater number of parameters estimated. Suitable choices of the value for κ eliminate models that are too simple to fit the data well along with unnecessarily general models that overfit the data. The best model will provide an account of the data that is both accurate and parsimonious.

Different model selection criteria use different values for κ. Theoretical considerations rule out values for κ that are too small or two large. *HQ* (Hannan & Quinn, 1979) sets $\kappa = 2\ln(\ln(N))$, where $\ln(N)$ is the natural logarithm of the sample size N. (The natural log is taken twice.) This results in $\kappa = 3.64$ for the 481 observations of our actual data set. *BIC* (Schwarz, 1978) uses $\kappa = \ln(N)$, which equals 6.18 for our actual data. Any value of κ between these two is equally defensible.

We chose to set $\kappa = 3.84$, which corresponds to estimating a parameter as part of the model if and only if the estimate's level of statistical significance is less than or equal to 0.05. We will denote this criterion as $MSC_{.05}$. There is value in using a criterion consistent with common statistical practice. Our choice of κ imposes a penalty intermediate to the *HQ* and *BIC* extremes.[5] It is closer to *HQ* than to *BIC*, which will eliminate parameters whose estimates have levels of statistical significance as low as 0.013.

We do *not* use the so-called "hit rate," which is a tally of the number of decision outcomes correctly fitted by a model. According to this criterion, a model is said to correctly fit a binary decision if the fitted probability of the observed decision is greater than one-half. This rule is undesirable for several reasons. (a) Adding parameters to a model can never be expected to lower the hit rate of the model *in sample*. Therefore, models assessed using hit rates must be estimated on only a portion of the data and then their hit rates calculated for the remaining portion. The result of this procedure depends upon how the data happen to be divided. (b) Since the hit rate scores all fitted probabilities in the interval (0.5, 1.0) as "correct," it has little ability to distinguish among models that differ widely in their accuracy.[6] Most importantly for our application, the hit rate does not distinguish a noncompensatory model predicting certain rejection of unacceptable objects from a compensatory model predicting these same objects have probabilities of acceptance as high as one-half.

The *MSC* criteria, given by (13), do not share in these shortcomings. These penalized likelihood measures take the fitted model probabilities into account along with the number of parameters used to obtain these fits. All available data are used both to estimate the model and to assess its performance.

Application to actual decisions

We illustrate the application of our additive model to actual MBA admission decisions. We have applied the model to graduate admissions data using expert decision makers because, as Phelps and Shanteau (1978) suggest, expert decision makers are better able to integrate multi attribute information. Thus, while the reliance on heuristics is not restricted to laymen (Tversky & Kahneman, 1974), the expert decision maker may be better able to trade off information in a compensatory manner and thus provide a strong test of our model. These data have the additional advantage that all decisions were made by only two individuals working in concert, so false conclusions about decision strategies due to the aggregation of data from heterogeneous individuals are unlikely.

The data consist of 481 records of applicants over a 2-year period. We modeled the binary dependent variable—whether each applicant was admitted—using the logit model of (11). Five independent variables were included in the analysis. Two of these were binary: year of application (YEAR) and a variable indicating whether the applicant received his or her undergraduate degree in the US or Canada rather than in some other country (US-CAN). The remaining three independent variables are intervally scaled and numerous different values occur in the data for each. These variables were the applicant's undergraduate grade point average, converted to the 1–9 grading scale employed by this university (GPA), score on the Graduate Management Achievement Test (GMAT), and years of full-time work experience in management and/or engineering (WORKME).

The two binary variables were included additively in the model without transformation. GMAT, GPA, and WORKME were standardized but were otherwise left untransformed. Based on our experience with admissions, we expected that departures from the linear model, if any, would be in the form of diminishing returns—that is, low scores on GPA, GMAT or WORKME would hurt an applicant more than high scores would help but, ceteris paribus, higher scores always remain preferable to lower ones. We also expected to see evidence of conjunctive or conjunctive-plus-linear behavior for some or all of these attributes.

Model estimates and statistical assessment

All three models plus the linear compensatory model were estimated using S-Plus (Insightful Corp., 2001). The polynomial and smoothing spline models are both

[5] This will not be true for every dataset, since the values for HQ and BIC both depend upon the sample size.

[6] The inability of the hit rate measure to distinguish among competing models was confirmed for our application to actual data, described next.

Table 2
Estimates of all additive models, actual MBA admit/reject decisions

Attribute	GNH[a]		Polynomial (cubic)[b]		Smoothing spline[a]		Linear[a]	
	Full[c]	Final[d]	Full[c]	Final[d]	Full[c]	Final[d]	Full[c]	Final[d]
Constant	−1.87	+6.85	−2.31	−1.93	−1.12	−1.10	−1.13	−1.12
YEAR[e]	−0.30		−0.33		−0.32		−0.27	
USCAN[e]	+0.78	+0.75	+0.69	+0.70	+0.71	+0.67	+0.70	+0.67
GMAT	$\beta = +1.79$	+2.88	$\beta_1 = +4.92$	+3.90	$\beta = +3.12$	$\beta = +3.04$	+3.18	+3.13
	$\gamma = +2.17$		$\beta_2 = -2.04$	−1.15	(−2 df^{f})	(−1 df^{f})		
	$\rho = -0.62$	−1.00	$\beta_3 = +0.87$					
GPA	$\beta = +0.00$		$\beta_1 = +1.96$	+1.95	$\beta = +1.72$	$\beta = +1.64$	+1.66	+1.61
	$\gamma = +2.04$	34.45	$\beta_2 = -0.70$	−0.73	(−2 df^{f})	(−1 df^{d})		
	$\rho = -0.24$	−0.23	$\beta_3 = -0.12$					
WORKME	$\beta = +0.00$	+1.25	$\beta_1 = +0.87$	+0.77	$\beta = +1.07$	$\beta = +0.93$	+0.88	+0.84
	$\gamma = +1.22$		$\beta_2 = -0.57$	−0.70	(−2 df^{f})	(−1 df^{f})		
	$\rho = -0.55$		$\beta_3 = +0.19$					
$MSC_{.05}$[g]	304.8	**292.5**	299.4	292.9	301.3	295.8	305.8	305.2

[a] GMAT, GPA, and WORKME were standardized.
[b] For GMAT, GPA and WORKME, the quadratic and cubic variables were created from orthogonal polynomials that were standardized.
[c] The most general form of the model estimated.
[d] The final form, containing only statistically significant parameters ($p < .05$).
[e] YEAR and USCAN were coded using dummy variables.
[f] Smoothing splines represent curvature nonparametrically. However the number of degrees of freedom lost can be calculated for any degree of curvature.
[g] Model Selection Criterion—a measure of model performance, with lower values indicating superior models. Models with too many or too few parameters will score poorly.

implemented in S-Plus's `gam` function, the linear compensatory model was estimated using S-Plus's `glm` function, and the GNH model was programmed in the S-Plus language.[7]

The first numerical column of Table 2 shows the estimates of the most general ("full") GNH model, with 12 estimated parameters. The column to its right shows the estimates of the "final" GNH model that resulted from a forward/backward stepwise modeling procedure, described below. The remaining columns in the table show the results for the full and final models for the polynomial, smoothing spline and linear regression models, respectively. The last row of Table 2 shows the model selection criterion $MSC_{.05}$ for all models.

The full versions of the GNH, cubic polynomial and smoothing spline models all estimated the same number of parameters. All three of these models are superior to the linear compensatory model. The full GNH model has estimates of ρ between −1 and 0 for all three attributes, implying a pervasive conjunctive plus linear rule for each. The $MSC_{.05}$ scores for the full polynomial and smoothing spline models are better than for the full GNH model, but both imply substantial nonmonotonic value functions for at least one attribute. Furthermore, all four full models contain numerous statistically nonsignificant parameters, indicating that the full generality of these models are overly complex characterizations of the decision process.

Given sufficient theoretical guidance and sufficient data, the initial model should be the only model estimated. However, these conditions are often not satisfied, as is the case here. Therefore, we implemented the same automated stepwise model selection procedure for each model to eliminate from each model parameters that are not warranted by the data. Improvement upon a full model was sought by comparing its $MSC_{.05}$ score to that of minor variations in the model. Of these models, the one with the best $MSC_{.05}$ score is then compared to *its* minor variants. The process is repeated until it results in no further improvement. For all models, the minor variants considered included all models that altered the treatment of any single one of the attributes. All treatments of any variable that entail estimation of one fewer, one more, or the same number of parameters were considered at each step.

We see from Table 2 that the stepwise procedure improved upon the full model in every case. All parameter estimates in every final model have levels of statistical significance less than 0.05. Every final model retains all variables except YEAR. The best model by

[7] Both `gam` and `glm` are also available in R, a public domain statistical language. See http://cran.R-project.org/. `gam` is to be found in the `mgcv` package distributed with R.

the $MSC_{.05}$ criterion is the final GNH model. It treats GPA and GMAT nonlinearly, estimating two parameters for each of them, but treats WORKME linearly. It is the best model of all shown in Table 2 according to the *HQ* and *BIC* criteria as well.

A model that performs nearly as well as the GNH according to $MSC_{.05}$ is the final polynomial model, which fits all three intervally scaled variables using quadratics. The final smoothing spline model performs less well, estimating the equivalent of one nonlinear parameter for each of the intervally scaled variables. The linear compensatory model performs worst of all.

Theoretical assessment: Examining the estimated value functions

While overall measures of model performance are of value, it is important to examine critically the decision rules implied by each model. We examine how the four final models of Table 2 treat each of the three numeric variables.

GMAT. The top portion of Fig. 3 shows the estimated relationship between the variable GMAT and its valuation (*V*) as estimated by the linear, quadratic, smoothing spline, and GNH final models. To facilitate comparisons, we adjusted the intercepts so that all four functional forms assign a utility of zero to the median GMAT score of 560.

The bold solid line in Fig. 3 is the fit of the crisp-conjunctive-plus-linear GNH value function. It implies that no applicant with a GMAT below 500 would be admitted to the program, and in fact none of the 90 applicants with GMATs below 500 (19% of the sample) were admitted. A minimum GMAT of 500 is plausible for this school, providing a degree of face validity. The mean GMAT for those admitted was 615. The bottom portion of the figure shows the frequency distribution for the observed GMAT scores for all applicants.

The quadratic curve illustrates the difficulty the polynomial value function has in capturing nonlinearity without introducing nonmonotonicity. The quadratic cannot quite match the decline of the GNH value function for low GMAT scores without increasing the decline in utility for GMAT scores above 715.

The nonmonotonicity of the quadratic curve in Fig. 3 is more marked than it appears. Since in a logit model each decline in valuation of 0.7 implies a halving of an applicant's odds of acceptance, the quadratic implies that an applicant with a GMAT score of 715 would be two and one-half times *more* likely to be admitted as one with a perfect GMAT score of 800, even when holding all other variables constant. Nonmonotonicity is not plausible for this attribute.

Finally, the smoothing spline fit indicates that some nonlinearity in the effect of GMAT is statistically justified. It shows a diminished effect of GMAT for high scores. Notice that the estimated curvature is all in the

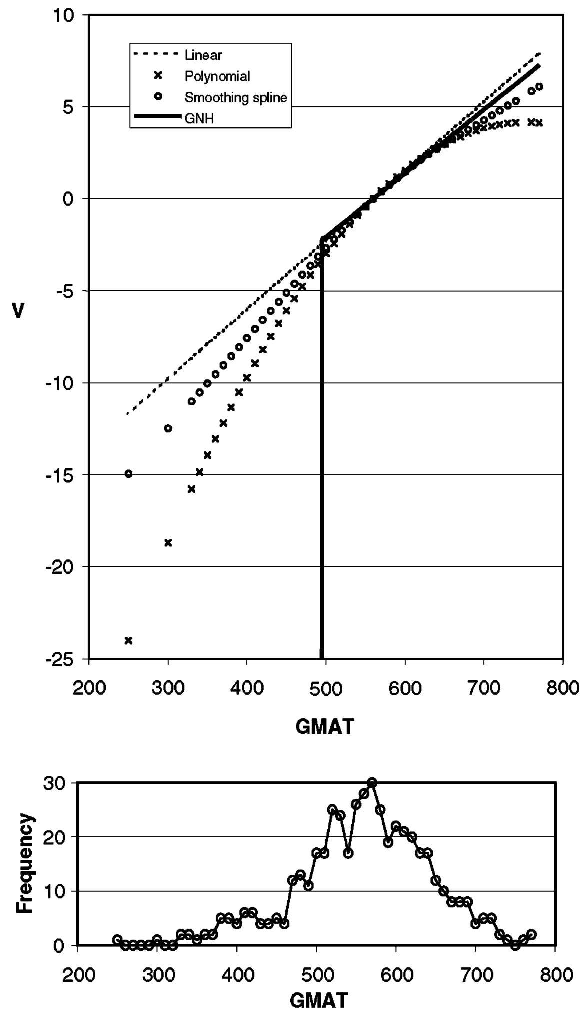

Fig. 3. Estimated value functions for GMAT, actual data.

region of GMAT scores with high frequency in the data. Smoothing splines have a bias towards assuming linearity in regions with few data points (Hastie & Tibshirani, 1990).

GPA. We see from Fig. 4 that the differences among models for the GMAT also hold for GPA, except that the GNH model fits a pervasive conjunctive form for GPA rather than crisp conjunctive plus linear. While the estimated cutoff value lies well within the range of possible GPA values, it lies below all observed GPA values, so no applicants were rejected outright on this attribute alone.

The final polynomial model fits a quadratic to this variable as well, and again implies a nonmonotonic relationship between the attractiveness of applicants and their GPAs over the range of the data. It can approximate the negative curvature for low values of GPA only by inducing nonmonotonicity for high values. The agreement between the GNH and smoothing spline models for high values of GPA is remarkably close. However, the smoothing spline model again returns to linearity for low values of GPA, where there are fewer

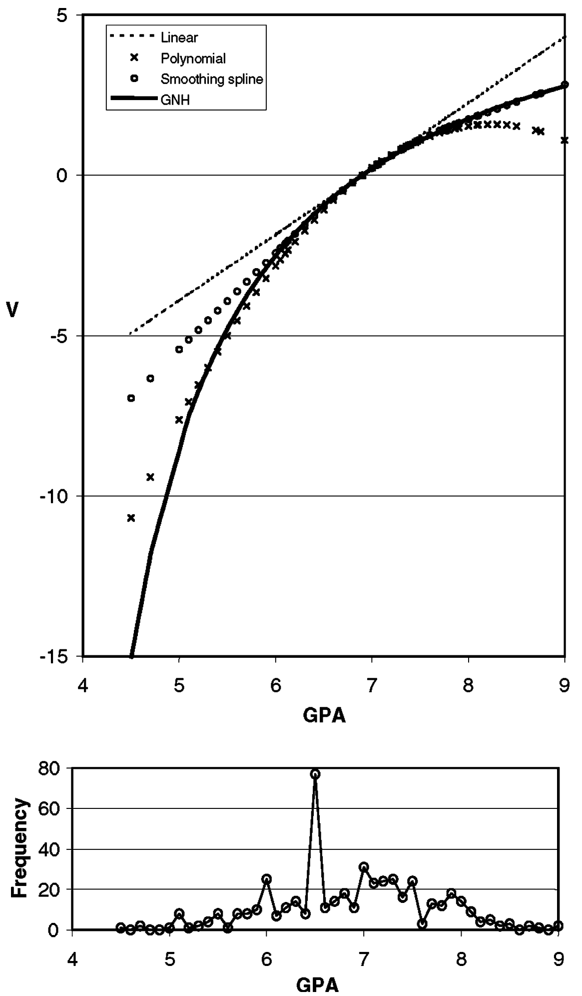

Fig. 4. Estimated value functions for GPA, actual data.

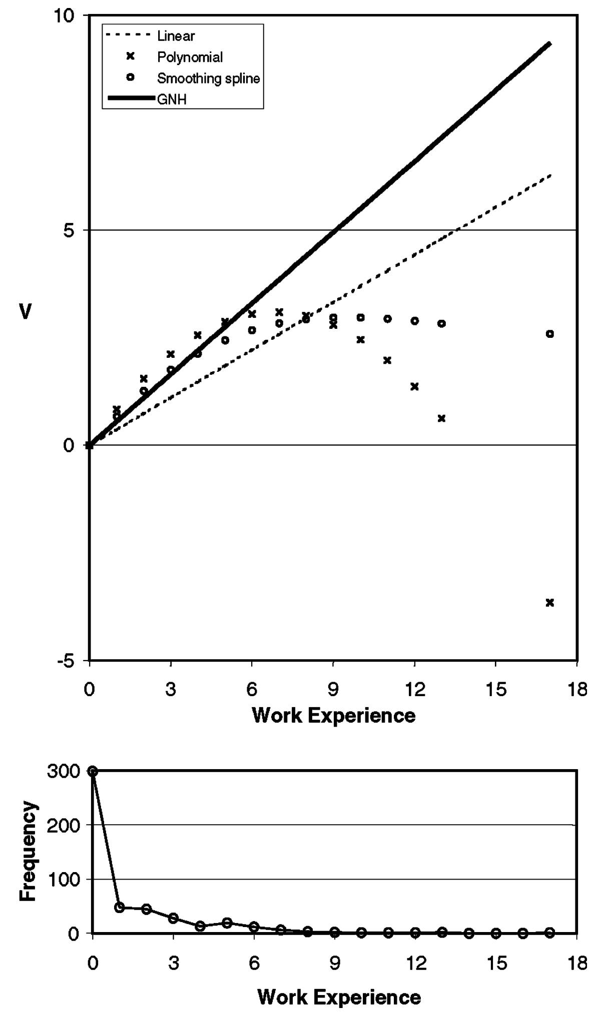

Fig. 5. Estimated value functions for work experience, actual data.

data points, whereas the GNH value function assumes that the diminishing effect of higher GPA holds over the entire range of the data.

WORKME. Some further insights into the models result from considering how they handle the WORKME variable (Fig. 5). Both the polynomial and smoothing spline models find a nonmonotonic relationship between work experience and acceptability of an applicant, even after controlling for the effect of all other applicant characteristics. The nonmonotonicity implied by the polynomial model is extreme—it implies that all applicants with eight or more years of work experience will be deemed less desirable than one with seven. It seems more likely that the polynomial model is again revealing its inability to model nonlinearity over one region of the data without inducing nonmonotonicity in other regions. The smoothing spline also implies nonmonotonicity, but to a much smaller extent.

Interestingly, the GNH model finds no statistically significant indication of nonlinearity in the treatment of WORKME. The detection of nonlinearity (and nonmonotonicity) for WORKME by the other two models is likely due to WORKME's relation to GMAT. Of the 11 applicants with more than 7 years of work experience, 45% (5 out of 11) had GMAT scores less than the estimated cutoff of 500, compared with only 18% of the other applicants, a difference that is marginally significant ($p = .056$). Since only the GNH model assigns zero probabilities of acceptance to all applicants with GMATs below 500, the other models may instead infer nonmonotonicity for WORKME because of its association with very low GMAT scores.

Overall assessment

The final GNH model provides the best characterization of the data according to a model selection criterion. Furthermore, it agrees better with our prior knowledge of the decision making process. Comparisons of models on real decisions are hampered by the inherent unknowability of the "true" decision rule employed, but the GNH result is more plausible than those of the alternative models. Examination of the estimated value functions of all final models reveals that the smoothing spline and polynomial models imply nonmonotonic valuation on one and two attributes, respectively, that are believed to be monotonically related to an applicant's merit. The GNH model estimated conjunctive cutoffs for two attributes. Its estimate of 500 as a minimally acceptable GMAT score is particularly plausible and is consistent with decision maker self-reports.

Application to simulated decisions

The analysis of actual admission decisions showed that the GNH model produces results that characterize the data well and are plausible in terms of theory, but it cannot establish the model's ability to capture the true, unknowable, decision strategy. To test further the GNH model's ability to capture specific decision strategies, we simulated ten different data sets that contained binary decisions resulting from different decision strategies applied to three attributes. We then compared the abilities of the GNH, polynomial and smoothing spline models to identify the strategies used. This procedure enables us to go beyond speculation about actual decision rules because the actual decision rules are known.

Simulation design

The simulated data were generated by the second author and presented to the first author for analysis. Values for the three attributes were generated for 1000 decisions using independent standard normal distributions. In addition, binary decisions were generated using ten different rules that involved some or all of the three attributes. The rules were implemented by a computer program (written in BASIC) in accordance with descriptions of the compensatory, conjunctive, disjunctive, and hybrid decision strategies in the J/DM literature (e.g., Billings & Marcus, 1983; Dawes, 1964) rather than in the manner assumed by any one of the GNH, polynomial or smoothing spline value functions. One dataset used the linear compensatory value function for all attributes, eight datasets treated some attributes in a noncompensatory manner (four as pervasive and four as crisp), and one included a nonmonotonic attribute.

Each of the simulated datasets was analyzed using the same model selection procedure and criterion as were used with the graduate admissions data. Even though this model fitting and selection process is completely automatic, the data analyst was nonetheless kept ignorant, at the time of analysis, of the strategies that were used to generate the data.

Statistical assessment

Fig. 6 portrays the performance of these models. Since the GNH performed best in all cases but one,

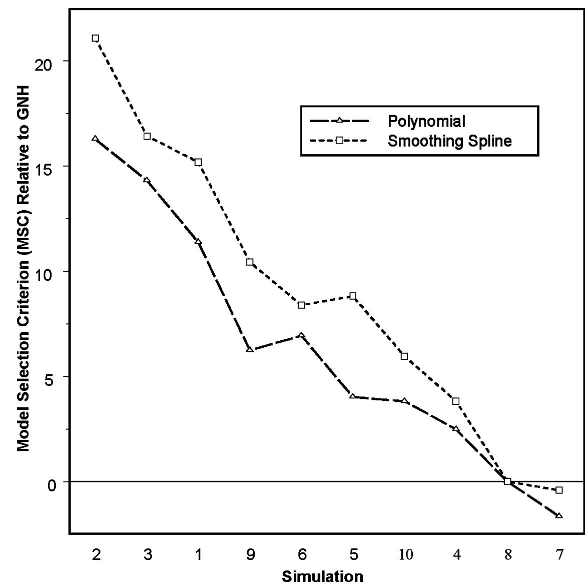

Fig. 6. Performance of two methods relative to the GNH for 10 simulated data sets.

Fig. 6 plots the difference between the $MSC_{.05}$ for the polynomial/smoothing spline models and the GNH model. Since lower $MSC_{.05}$ scores are better, most of these differences are positive. These differences are sorted from largest to smallest to improve readability.

The GNH model performed better than the other two models for eight of the 10 data sets. The three models performed identically for simulation 8 because the true model was linear in all three attributes and all three models correctly identified this model. The GNH performed worse only for simulation 7.

The true model for simulation 7 was quadratic in the third attribute and nonmonotonic over the range of the data. This explains why the polynomial model performed best with these data—the true model is polynomial and the stepwise modeling procedure applied to the polynomial case correctly identified the true model. This also explains why the best smoothing spline model performed better than the GNH—it can and did identify the nonmonotonic effect of the third attribute. This example is useful because it shows that the GNH model does not automatically perform better when the data depart from the model's behavioral foundations. While extant theory and empirical evidence suggest that monotonic evaluation of attributes predominates in field settings, researchers are free to adopt a nonmonotonic value function such as the polynomial for some attributes, when deemed appropriate, and the GNH for others. Therefore, remaining discussion focuses on the other nine simulations, which preserve monotonicity for all attributes, and considers the adequacy of polynomials and smoothing splines relative to the GNH in these settings.

Theoretical assessment: Model differences in attribute characterizations

Each model must determine, for each simulation, how each of the three attributes affects the response; specifically, all three models can represent an attribute effect as either absent, linear or nonlinear. Given nine simulations and three attributes, each model must therefore make 27 such determinations. The GNH model made only one incorrect categorization out of 27, whereas the polynomial and smoothing spline models each made three.

The nine simulations included 18 noncompensatory attribute effects, nine crisp and nine pervasive. The GNH model correctly classified 15 of these as crisp or pervasive, making three errors. The smoothing spline and polynomial models cannot identify crisp noncompensatory rules but can only approximate them using nonlinear compensatory functions; thus, they each necessarily made nine errors.

Finally, all attribute effects in the nine simulations were monotonic. While the GNH value function imposes monotonicity for the effect of an attribute, the smoothing spline and polynomial models cannot. Each of these two models inferred nonlinearity in 20 instances. The smoothing spline model *incorrectly* inferred nonmonotonicity over the range of the data in 60% of these cases, and the polynomial model in 80%.

The frequent failure of the smoothing spline model to estimate monotonicity over the range of the data is surprising. It should tend to preserve monotonicity because it penalizes curvature, and nonmonotonic models are more curved. The frequent findings of nonmonotonicity in this case may be a result of the numerical procedure used in estimation.[8] The polynomial model's inherent difficulty in fitting nonlinearity without violating monotonicity was expected, but the simulation demonstrates the extreme seriousness of this limitation.

Discussion

In this section, we describe potential uses of the GNH model in its present form, show how it might be implemented when studying sequential or multiple choice tasks, and then consider promising extensions of the model.

Investigating substantive issues

By enabling J/DM researchers to identify decision strategies from outcome data, the GNH model eliminates the need to study decision makers in a laboratory rather than in the midst of the distractions and incentives of the real world. The ability to study the prevalence of various decision strategies under naturalistic conditions would constitute an important extension of the literature on strategy selection (e.g., Broeder, 2003; Payne et al., 1993).

The GNH model has been investigated in detail here only for independent binary decision making. Even within this context, the model may be used to answer interesting substantive questions.

How prevalent are each of the nine rules listed in Table 1? How does choice of rule vary with the characteristics of the task and of the decision maker? The GNH model can contribute to the substantial literature regarding this question by enabling researchers to examine actual choices made outside the laboratory.

Do decision makers tend to use the same rule on all attributes, or does rule usage vary systematically with attribute characteristics? We have seen that some JD/M researchers regard decision strategies as being object-specific but not attribute-specific. For example, decision makers using a conjunctive strategy are assumed to employ the conjunctive rule for all attributes. The GNH model may be used to test this assumption, since it allows decision makers to process different attributes differently.

Alternatively, rule usage may vary by attribute in unpredictable ways. For example, decision makers may use cognitively more demanding rules for the more important attributes (i.e., those attributes that explain most of the variability in their decisions). We would expect that the complexity of the rules of Table 1 adhere to the partial ordering: {crisp conjunctive/disjunctive} < {linear, pervasive conjunctive/disjunctive, and crisp conjunctive/disjunctive plus linear} < {pervasive conjunctive/disjunctive plus linear}, but empirical investigation is required.

Modeling sequential decision making

In sequential decision making, the decision maker examines objects one at a time until an acceptable object is found and selected. The researcher observes the attributes of all N_t objects that were examined, and that the N_tth one was chosen. Then we know that

$$U_{ti} = \alpha + \sum_{q=1}^{Q} V_{tiq} + \varepsilon_{ti} \tag{14}$$

is negative for $i = 1, \ldots, N_t - 1$ and positive for $i = N_t$. Note that the intercept α is included for all examined objects. Reliable estimation of any multi attribute choice model, including the GNH, would require multiple (i.e., $t = 1, \ldots, T$) decisions of this nature.

Modeling multiple choice tasks

Sometimes a decision maker must choose one alternative from a set of N.[9] In such cases, we observe a

[8] The `gam` function in S-Plus uses an iterative "local scoring" estimation procedure for smoothing splines that compromises precision in order to reduce estimation time (Hastie & Tibshirani, 1990).

[9] N may vary from one multiple choice set to the next, i.e. we may replace N with N_t, provided that it is exogenously determined.

single choice from each of T choice sets, $t = 1, \dots, T$. The utility of alternative i in choice set t may be denoted U_{ti}, and $Y_t = i$ if and only if $U_{ti} = \max\{U_{t1}, \dots, U_{tN}\}$. The random utility model of (1) is readily adapted to this case. Utility is now represented as $U_{ti} = V^*_{ti} + \varepsilon_{ti}$, where V^*_{ti} represents the deterministic utility of alternative i in choice set t. We briefly discuss two different types of multiple choice tasks: *forced choice* of an object, where all N alternatives in the choice set are objects, and *optional choice* of an object, where one of the alternatives is "choose none", which allows no choice of any of the $N-1$ objects in the choice set.

Forced choice of an object

If all of the alternatives are objects, then decision makers are forced to always choose an object from each set. Deterministic utility for all alternatives is given by

$$V^*_{ti} = \sum_{q=1}^{Q} V_{tiq}, \quad i = 1, \dots, N, \tag{15}$$

where $V_{tiq} \equiv f_q(X_{tiq})$ represents the possibly nonlinear value function for attribute q of object i in choice set t. Notice that the intercept α found in (1) is missing from (15). This is necessarily so because choice from a set of objects provides information only about the relative utility of the objects. Notice also that the value function $f_q(X_{tiq})$ may differ across attributes but it remains the same for all objects and choice sets. The result of the function will change as X_{tiq} changes, but only one function must be estimated for each attribute. When there are only two objects in the choice set, then the probability of the second being chosen simplifies to

$$Y_t = 2 \text{ if and only if}$$
$$U_t = (V^*_{t2} - V^*_{t1}) + (\varepsilon_{t2} - \varepsilon_{t1}) \equiv V^*_t + \varepsilon_t > 0. \tag{16}$$

Optional choice of an object

We can avoid forced choice of an object by making the first alternative "choose none." Decision makers still make a forced choice from the set of N *alternatives* but may avoid choice of an unacceptable *object* by choosing alternative 1. Therefore this case may still be represented as $Y_t = i$ if and only if $U_{ti} = \max\{U_{t1}, \dots, U_{tN}\}$. Y_t will equal one whenever the utilities of all $N-1$ objects in the choice set are lower than the minimum acceptable utility.

The deterministic utility of "choose none" does not arise from evaluation of the attributes of this alternative—it has none—but is simply an unknown constant. This unknown constant may be denoted $-\alpha$ or, equivalently, the deterministic utility of the first alternative may be fixed at zero and an intercept α may be included in the deterministic utilities of all $N-1$ objects in the choice set. This latter approach results in

$$\begin{aligned} U_{t1} &= 0 + \varepsilon_{t1}, \\ U_{ti} &= \alpha + \sum_{q=1}^{Q} V_{tiq} + \varepsilon_{ti}, \quad i = 2, \dots, N. \end{aligned} \tag{17}$$

When $N = 2$ we have unforced choice of a single object and (17) simplifies further to give

$$\begin{aligned} Y_t = 2 \text{ if and only if } U_t &= (V^*_{t2} - V^*_{t1}) + (\varepsilon_{t2} - \varepsilon_{t1}) \\ &\equiv V^*_t + \varepsilon_t \equiv \alpha + \sum_{q=1}^{Q} V_{tq} > 0. \end{aligned} \tag{18}$$

Since there is only one real object in each choice set, the distinction between object and choice set disappears, and $V_{t2q} \equiv V_{tq}$.

Error term distributions

Random utility models are not consistent with all possible observed outcomes. While the constraints imposed on choice probabilities *inherent* to random utility models—nonnegativity of the Block–Marschak functions (Block & Marschak, 1960; Falmagne, 1978)—are relatively weak, particularly for binary choices (Iverson & Luce, 1998), they are still violated in some situations, and the choice of distribution for the random error terms leads to additional restrictions.

The popular (multinomial) logit model is the most restrictive in common use, implying independence from irrelevant alternatives and strong stochastic transitivity. Numerous theoretical and empirical studies have shown that these properties do not hold in some settings (cf. Busemeyer & Diederich, 2002; Busemeyer & Townsend, 1993). More complex treatments of the error terms in the multinomial case are possible and, in many settings, desirable.

We note two common choices of distributions for the error terms when modeling multiple choice. The simplest is to assume that the errors $\varepsilon_{t1}, \dots, \varepsilon_{tN}$ are independently and identically distributed according to the double-exponential distribution.[10] This results in the (conditional) *multinomial logit* model, with

$$\Pr(Y_t = i) = \exp(V^*_{ti}) \Bigg/ \sum_{i'=1}^{N} \exp(V^*_{ti'}). \tag{19}$$

When $N = 2$, the error term ε_t in (16) and (18) has the logistic distribution. When $N = 2$ and choice of object is optional, then the probability of choice is given by (11).

The multinomial logit model assumes that the random error terms for the different objects in the choice set are independently and identically distributed. A natural alternative is the multinomial probit, which allows the error terms to have a general multivariate normal distribution that is estimated from the data.[11] It implies the

[10] $\Pr(\varepsilon \leqslant \varepsilon_0) = \exp[-\exp(-\varepsilon_0)]$.

[11] A variance parameter must be fixed, without loss of generality, in order to identify probit models.

less stringent properties of moderate stochastic transitivity and regularity (Fishburn, 1998; Busemeyer & Diederich, 2002; Halff, 1976).

When $N = 2$, the multinomial probit model reduces to the simple probit model, which is empirically indistinguishable from the logit model except in exteremely large samples (Chambers & Cox, 1967). However, repeated choices from more than two alternatives often reveal the inflexibility of the multinomial logit model, in which case a multinomial probit specification for the GNH model would be worthwhile.

Noncompensatory rules and multiple choice tasks

Justification for noncompensatory rules such as the conjunctive has generally been offered in two predominant settings: in sequential decision making where the first acceptable object is chosen, and in two-stage decision making in which a noncompensatory rule is used to reduce the set of objects to be considered more carefully. In multiple choice settings, however, one and only one alternative can be selected, and it is worth considering what sorts of rules ought to apply in such settings.

Clearly, straightforward application of deterministic conjunctive or disjunctive rules will not automatically yield exactly one selected alternative from every imaginable choice set, and strict application of noncompensatory rules may be seen less often in such settings. We offer here some conjectures about the kinds of rules that might be expected, and indicate how they would be represented by the GNH model. We consider both optional and forced choice of an object in a multiple choice task.

Optional choice of an object using a rule with conjunctive characteristics

In this task, the decision maker may choose either zero or one object from a choice set. Here we consider the GNH model with conjunctive characteristics, i.e., the value functions generated by (10) with $-1 \leqslant \rho \leqslant 0$ illustrated in Fig. 1. Conjunctive elimination of all objects is a viable result for this task, but selection of more than one object must be avoided.

How might a decision maker choose among two or more objects that meet a noncompensatory criterion? Three possibilities are represented by the GNH value function exactly: (a) the remaining objects are evaluated in a linear compensatory manner (resulting in the conjunctive-plus-linear rule); (b) the object that most decisively meets the conjunctive criterion is chosen (pervasive conjunctive rule); and (c) choice among the remaining alternatives is random (conjunctive plus random error rule). In fact, the GNH model handles any combination of these three possibilities exactly. A fourth possibility is that choice is made in a sequential manner but the researcher does not know which objects were examined and rejected and which were simply never examined. In this case, which of the numerous acceptable objects happened to be examined first and hence chosen is, from the researcher's perspective, random and can be modeled as such (as in (c)).

Forced choice of an object using a rule with conjunctive characteristics

Here choice of an object from the set is required. The GNH model can represent conjunctive evaluation in forced choice of an object as pervasive conjunctive or pervasive conjunctive plus linear. Either of these models may be deterministic or may contain random error.

Using a rule with disjunctive characteristics

This case applies to the GNH model with disjunctive characteristics, i.e., the value functions generated by (10) with $0 \leqslant \rho \leqslant +1$ illustrated in Fig. 2. Regardless of whether choice of an object is optional or forced, both the pervasive disjunctive and pervasive disjunctive-plus-linear value functions can account for observed choices matching the task requirements, either with or without random error.

Bayesian estimation

Estimating all parameters of flexible value functions (such as the three-parameter GNH) for all attributes by maximum likelihood can be expected to result in some uncertain estimates. (This problem arose with all full models estimated on the actual data set.) Two techniques are available to deal with this problem.

1. Selected parameters may be fixed at known values (as shown in the third column of Table 1). We have illustrated the use of a model selection criterion to identify statistically insignificant parameters that may be fixed at known values without appreciable loss of model fit.
2. Equality constraints may be imposed on functions of subsets of the parameters. The identity constraint is most often used. For example, we might force the estimated value of ρ_q of (10) to be equal across all Q attributes. Note that the first technique is a special case of this one.

Bayesian estimation (Denison, Holmes, Mallick, & Smith, 2002) allows use of an additional tool for stabilizing the estimates of models containing many parameters—distributions may be estimated for any subsets of parameters. For example, we might assume that the values of $\rho_1, \ldots, \rho_Q$ for the Q different attributes are drawn from a some common distribution, and estimate the parameters of this distribution. Typically, the estimated distributions of parameters will be unimodal, which implies that the parameters themselves (in this example, $\rho_1, \ldots, \rho_Q$) are neither completely unrelated nor exactly equal. Alternatively, predicted values for the parameters might be an estimated function of observed characteristics of the attributes, in which case the estimated common distribution would apply to random departures of the parameters from their predicted values.

Bayesian estimation does have drawbacks which should be acknowledged. It requires specialized software and in-depth knowledge of statistics. Estimation is typically slow, which hampers model selection, and classical hypothesis tests are unavailable.

Modeling unexplained heterogeneity in decision making

In many settings, we wish to understand decisions by numerous decision makers who are not necessarily acting in concert. In other settings, the same decision makers may use somewhat different rules over time. In either case, it is straightforward to allow parameters to be estimated functions of observed characteristics of decision makers and/or of the decision context. Allowing GNH model parameters to also vary *randomly* across decision makers or decision contexts (reflecting the influence of unobserved decision maker characteristics) is practicable, especially with Bayesian estimation (Denison et al., 2002). The result may be viewed as an extension of the wandering vector model (Carroll, 1980, Carroll & De Soete, 1991).

Incorporation into decision field theory

In the Introduction, we briefly mentioned use of connectionist models by J/DM researchers. Developers of decision field theory (Busemeyer & Diederich, 2002; Busemeyer & Townsend, 1993; Roe et al., 2001) have used specialized connectionist models for the decision making process that reflect a growing understanding of the operation of the human mind. In particular, this literature includes implementation of dynamic stochastic models of evaluation and choice. For each choice occasion, the assessment of the alternatives varies randomly about expected values that depend upon object attributes, immediately previous assessments and also, possibly, time trends. These models may be estimated from information on choices and optionally on decision times (Diederich & Busemeyer, 2003). They have been shown to explain in plausible terms previously unexplained violations of common assumptions of choice models such as strong stochastic transitivity and regularity.

To date, applications of decision field theory have used the linear compensatory model to combine object attributes into object assessments, and the probabilistic formulation used by these models has precluded any object from being rejected or accepted with certainty as the result of application of a noncompensatory rule. Incorporation of the GNH model could relax these assumptions while retaining the virtues of the decision field theory. The resultant model, while very general, need not always be estimated in its most general form.

Comparison to Gilbride and Allenby (in press)

A forthcoming paper (Gilbride & Allenby, in press) estimates models of hybrid noncompensatory-compensatory decision making from experimental multiple choice data. Gilbride and Allenby compare the linear compensatory model and four hybrid models. Only one of their hybrid models can represent decisions as resulting from object valuation using a single additive multi attribute utility function (cf. (2)). This model—equivalent to a special case of the GNH model in which a crisp-conjunctive plus linear value function is applied to all attributes—performed best of all.

Gilbride and Allenby assume that a decision maker will use the same decision strategy (e.g., crisp-conjunctive plus linear) for all attributes, and they do not allow for a pervasive effect of cutoffs. However, since they use Bayesian simulation to estimate their model, they can and do allow the conjunctive cutoffs and linear coefficients to vary from one decision maker to the next. As noted in the Bayesian estimation section, this important advantage of Bayesian estimation extends to the GNH model.

Conclusion

We have proposed and tested a new model of decision making that enables us to identify compensatory, conjunctive, disjunctive, and hybrid decision strategies from naturally occurring outcome data, thus eliminating the need for self-reports or process data (including multiple observations of intermediate steps in the decision process a la Levin and Jasper, 1995). This is a significant advantage for researchers who do not have access to process or self-report data, and for those who are concerned about the accuracy of such data. Furthermore, this model allows for decision strategies to differ across attributes, in contrast to most previously published approaches.

Our empirical tests compared the new model to two extant additive models applicable to decision making—polynomials and smoothing splines—that share some of these advantages; i.e., they may be estimated solely from object descriptions and decision/evaluation outcomes and they allow for different decision rules for different attributes. Furthermore, none of these models assumes that decision makers evaluate more than one attribute at a time, making them more consistent with theory than nonadditive noncompensatory models (e.g., Brannick & Brannick, 1989; Ganzach & Czaczkes, 1995).

However, these two extant additive models can only approximate noncompensatory rules using nonlinear compensatory functions (i.e., they cannot identify noncompensatory cutoffs), and they cannot ensure that attribute value functions are monotonic. The general nonrectangular hyperbola (GNH) value function used here embodies noncompensatory and compensatory rules as special cases and, unlike the other two models, it can detect the sharp discontinuities of attribute evaluation implied by noncompensatory models while ensuring monotonicity in the evaluation of attributes. Furthermore, the GNH model allows noncompensatory cutoffs to have a pervasive effect.

These three additive models are compared on their fits to and characterizations of one actual and 10 simulated data sets. All models outperformed the linear model, even after taking into account the extra parameters that must be estimated. The GNH model is better suited to the data than the polynomial or smoothing spline models, which appeared to underfit nonlinearity in attribute evaluation where it arose and frequently inferred nonmonotonicity when it was absent. Furthermore, the GNH model is more consistent with J/DM theory and provides a more plausible account of behavior compared to the other models.

In addition, we find some empirical support for a pervasive effect of noncompensatory cutoffs on decision making. This finding is consistent with introspective reports and some empirical evidence based on self-reported cutoffs, but to our knowledge it is the first demonstration of pervasive cutoffs that is based on observed choices and is not potentially confounded by asking decision makers about their cutoffs before they make their decisions.

We investigate, and demonstrate the usefulness of, the GNH model using a simple random utility specification. Because the GNH value function subsumes linear compensatory and two noncompensatory rules, it might be usefully introduced into other types of models that presently make more restrictive assumptions about how object attributes are evaluated to produce object assessments but less restrictive assumptions about how object assessments affect decisions.

References

Abe, M. (1999). A generalized additive model for discrete-choice data. *Journal of Business and Economic Statistics, 17*, 271–284.

Beach, L. R. (1993). Broadening the definition of decision making: The role of prechoice screening of options. *Psychological Science, 4*, 215–220.

Billings, R., & Marcus, S. (1983). Measures of linear and nonlinear models of decision behavior: Process tracing versus policy capturing. *Organizational Behavior and Human Performance, 31*, 331–351.

Block, H. D., & Marschak, J. (1960). Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, H. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics* (pp. 97–132). Stanford, CA: Stanford University Press.

Borooah, V. K. (2002). *Logit and probit: Ordered and multinomial models*. London: Sage Publications.

Brannick, M. T. (1986). The development and evaluation of some models for detecting the presence of nonlinear combinations in phenomena of interest to industrial and organizational psychologists. *Dissertation Abstracts International, 47*, 3564B.

Brannick, M. T., & Brannick, J. P. (1989). Nonlinear and noncompensatory processes in performance evaluation. *Organizational Behavior and Human Decision Processes, 44*, 97–122.

Broeder, A. (2003). Decision making with the adaptive toolbox: Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 611–625.

Brucks, M. (1985). The effects of product class knowledge on information search behavior. *Journal of Consumer Research, 12*, 1–16.

Brunswik, E. (1940). Thing consistency as measured by correlation coefficients. *Psychological Review, 47*, 69–78.

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences, 43*, 345–370.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*, 432–459.

Carroll, J. D. (1980). Models and methods for multidimensional analysis of preferential choice (or other dominance) data. In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice* (pp. 234–289). Bern, Switzerland: Huber.

Carroll, J. D., & De Soete, G. (1991). Toward a new paradigm for the study of multiattribute choice behavior. *American Psychologist, 46*, 342–351.

Chambers, E. A., & Cox, D. R. (1967). Discrimination between alternative binary response models. *Biometrika, 54*, 573–578.

Cherkassky, V., & Mulier, F. (1998). *Learning from data: Concepts, theory and methods*. New York: Wiley.

Coombs, C. H., & Kao, R. C., (1955). Nonmetric factor analysis. Ann Arbor: University of Michigan Engineering Research Institute Bulletin No. 38.

Dawes, R. M. (1964). Social selection based on multidimensional criteria. *Journal of Abnormal and Social Psychology, 68*, 104–109.

Debreu, G. (1954). Representations of a preference ordering by a numerical function. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes* (pp. 159–165). Princeton, NJ: Princeton University Press.

Denison, D. G. T., Holmes, C. C., Mallick, B. K., & Smith, A. F. M. (2002). *Bayesian methods for nonlinear classification and regression*. Chichester, West Sussex, England: Wiley.

Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology, 47*, 304–322.

Dillon, S., (1998). *Descriptive decision making: Comparing theory with practice*. Unpublished manuscript, Department of Management Systems, University of Waikato, New Zealand.

Edwards, W.& Tversky, A. (Eds.). (1967). *Decision making*. Baltimore, MD: Penguin.

Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin, 73*, 221–230.

Einhorn, H. J. (1971). Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance, 6*, 1–27.

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review, 86*, 464–485.

Falmagne, J.-C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology, 18*, 52–72.

Fishburn, P. C. (1998). Stochastic utility. In S. Barbera, P. J. Hammond, & C. Seidl (Eds.), *Handbook of utility theory* (Vol. 1, pp. 273–318). Dordrecht: Kluwer.

Ford, J. K., Schmitt, N., Schechtman, S. L., Hults, B. M., & Doherty, M. L. (1989). Process tracing methods: Contributions, problems, and neglected research issues. *Organizational Behavior and Human Decision Processes, 43*, 75–117.

Ganzach, Y., & Czaczkes, B. (1995). On detecting nonlinear noncompensatory judgment strategies: Comparison of alternative regression models. *Organizational Behavior and Human Decision Processes, 61*, 168–176.

Gensch, D. H. (1987). A two-stage disaggregate attribute choice model. *Marketing Science, 6*, 223–239.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world*. New York: Oxford University Press.

Gilbride, T. J. & Allenby, G. M. (in press). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*.

Goldberg, L. R. (1971). Five models of clinical judgment: A clinical comparison between linear and nonlinear representations of the human inference process. *Organizational Behavior and Human Performance, 6*, 458–479.

Green, P., Krieger, A., & Bansal, P. (1988). Completely unacceptable levels in conjoint analysis: A cautionary note. *Journal of Marketing Research, 25*, 293–300.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 87*, 1–51.

Grossberg, S., & Gutowski, W. (1987). Neural dynamics of decision making under risk: Affective balance and cognitive–emotional interactions. *Psychological Review, 94*, 300–318.

Halff, H. M. (1976). Choice theories for differentially comparable alternatives. *Journal of Mathematical Psychology, 14*, 244–246.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B, 41*, 190–195.

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. Hall, New York: Chapman.

Huber, J., & Klein, N. (1991). Adapting cutoffs to the choice environment: The effects of attribute correlation and reliability. *Journal of Consumer Research, 18*, 346–357.

Insightful Corp. (2001). *S-Plus 6 for Windows guide to statistics, volume 1*. Seattle, WA: Insightful Corp.

Iverson, G., & Luce, R. D. (1998). The representational measurement approach to psychophysical and judgmental problems. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making* (pp. 1–79). New York: Academic Press.

Johnson, R. M. (1987). Adaptive conjoint analysis. In *Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*. Sun Valley, ID, 253–266.

Klein, N. M., & Bither, S. W. (1987). An investigation of utility-directed cut-off selection. *Journal of Consumer Research, 14*, 240–256.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*, 132–157.

Leven, S. J., & Levine, D. S. (1996). Multiattribute decision making in context: A dynamic neural network methodology. *Cognitive Science, 20*, 271–299.

Levin, I. P., & Jasper, J. D. (1995). Phased narrowing: A new process tracing method for decision making. *Organizational Behavior and Human Decision Processes, 64*, 1–8.

Luce, R. D. (1956). Semi-order and a theory of utility discrimination. *Econometrica, 24*, 178–191.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

McFadden, D. (1987). Econometric models of probabilistic choice. In C. F. Manski & D. McFadden (Eds.), *Structural analysis of discrete data*. Cambridge, MA: MIT Press.

Mela, C. F., & Lehmann, D. R. (1995). Using fuzzy set-theoretic techniques to identify preference rules from interactions in the linear-model—an empirical-study. *Fuzzy Sets and Systems, 71*, 165–181.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy, 59*, 371–404.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 7*, 231–259.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance, 16*, 366–387.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 534–552.

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

Payne, J. W., Bettman, J. R., & Luce, M. F. (1998). Behavioral decision research: An overview. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making* (pp. 303–359). New York: Academic Press.

Phelps, R. H., & Shanteau, J. (1978). Livestock judges: How much information can an expert use? *Organizational Behavior and Human Performance, 21*, 209–219.

Pras, B., & Summers, J. (1975). A comparison of linear and nonlinear evaluation process models. *Journal of Marketing Research, 12*, 276–281.

Ratkowsky, D. A. (1990). *Handbook of nonlinear regression models*. New York: Dekker.

Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge: Cambridge University Press.

Roberts, J. (1988). A grounded model of consideration set size and composition. *Advances in Consumer Research, 15*, 749–757.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review, 108*, 370–392.

Ross, G. J. S. (1990). *Nonlinear estimation*. New York: Springer-Verlag.

Schoenberg, I. J. (1946). Contributions to the problem of approximation of equidistant data by analytical functions, part a: On the problem of smoothing or graduation. *Quarterly of Applied Mathematics, 4*, 45–99.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Schoemaker, P. J. H. (1980). *Experiments on decisions under risk: The expected utility theorem*. Boston: Martinus Nijhoff Publishing.

Shugan, S. (1980). The cost of thinking. *Journal of Consumer Research, 7*, 99–111.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99–118.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6*, 649–744.

Swait, J. (2001). A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B, 35*, 903–928.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76*, 31–48.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review, 76*, 31–48.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Yntema, D. B., & Torgerson, W. S. (1961). Man–computer cooperation in decisions requiring common sense. *IRE Transactions of the Professional Group on Human Factors in Electronics, HFE-2*, 20–26.