Information Systems Research
Vol. 20, No. 4, December 2009, pp. 527–546
ISSN 1047-7047 | EISSN 1526-5536 | 09 | 2004 | 0527

informs®

DOI 10.1287/isre.1080.0198

How Incorporating Feedback Mechanisms in a DSS Affects DSS Evaluations

Ujwal Kayande
College of Business and Economics, Australian National University, Canberra, ACT 0200, Australia,
ujwal.kayande@anu.edu.au

Arnaud De Bruyn
ESSEC Business School, 95000 Cergy-Pontoise, France, debruyn@essec.fr

Gary L. Lilien, Arvind Rangaswamy
Smeal College of Business, Pennsylvania State University, University Park, Pennsylvania 16802
{glilien@psu.edu, arvindr@psu.edu}

Gerrit H. van Bruggen
Rotterdam School of Management, Erasmus University, 3000 DR Rotterdam, The Netherlands,
gbruggen@rsm.nl

Model-based decision support systems (DSS) improve performance in many contexts that are data-rich, uncertain, and require repetitive decisions. But such DSS are often not designed to help users understand and internalize the underlying factors driving DSS recommendations. Users then feel uncertain about DSS recommendations, leading them to possibly avoid using the system. We argue that a DSS must be designed to induce an alignment of a decision maker's mental model with the decision model embedded in the DSS. Such an alignment requires effort from the decision maker *and* guidance from the DSS. We experimentally evaluate two DSS design characteristics that facilitate such alignment: (i) feedback on the *upside potential* for performance improvement and (ii) feedback on *corrective actions* to improve decisions. We show that, *in tandem*, these two types of DSS feedback induce decision makers to align their mental models with the decision model, a process we call deep learning, whereas individually these two types of feedback have little effect on deep learning. We also show that deep learning, in turn, improves user evaluations of the DSS. We discuss how our findings could lead to DSS design improvements and better returns on DSS investments.

Key words: decision support systems; DSS design; feedback; learning; mental models; evaluations
History: Ritu Agarwal, Senior Editor; Giri Kumar Tayi, Associate Editor. This paper was received on February 20, 2008, and was with the authors $8\frac{1}{2}$ months for 2 revisions. Published online in *Articles in Advance* December 18, 2008.

1. Introduction

Technological and modeling advances have dramatically increased the availability and quality of model-based decision support systems (DSS) (Shim et al. 2002, Banker and Kauffman 2004). Many such systems (e.g., customer relationship management systems, retail marketing mix DSS, employee scheduling DSS, clinical prescription DSS, etc.) are designed to assist decision makers in environments in which: (i) the data available to aid decision making are voluminous and beyond human information processing capabilities, (ii) the link between decisions and outcomes is probabilistic or uncertain, and (iii) the decisions are repetitive. In such environments, it is highly unlikely that decision makers can consistently outperform recommendations from even a simple model-based DSS (Hoch and Schkade 1996). Yet, Umanath and Vessey (1995, p. 796) observe that "since human decision makers do not know the rationale behind the suggested recommendation, they are typically skeptical of the output produced and are therefore reluctant to use such systems." We examine whether a DSS will be perceived as more valuable if it enables users to internalize the rationale behind those recommendations. We use our findings to develop insights on how DSS should be designed to enable such internalization.

There are many well recognized examples of user resistance to (objectively good) model-based DSS in data-rich, uncertain environments. Managers in retail grocery chains must set prices daily for thousands of products, integrating information about retail price elasticities amid uncertain competitive reactions. Retail pricing DSS that include price-optimization models have been shown to dramatically outperform retail managers (Reda 2003, Montgomery 2005). Yet, Sullivan (2005) reports that only 5% to 6% of retailers use such DSS, with most managers preferring to use gut feeling for pricing decisions. Similarly, clinical DSS significantly improve clinical performance in prescribing decisions (Hunt et al. 1998), yet medical professionals are largely unwilling to use them (Sintchenko et al. 2004, Lai et al. 2006). Ashton (1991), Singh and Singh (1997), and Sieck and Arkes (2005) among others have noted decision makers' disinclination to use DSS in a variety of different environments, even when the models embedded in the systems are known to improve decision quality and performance. Several researchers have suggested that a lack of user understanding of the logic underlying DSS output leads to poor perceptions of the value such model-based DSS offer, leading to user resistance and impeded system use (e.g., McIntyre 1982, Davis 1989, Van Bruggen et al. 1996, Lilien et al. 2004). In the context of retail pricing DSS, Montgomery (2005, p. 375) suggests that "Unless the model can provide some intuition in understanding why this new strategy is better, users are more apt to reject it." Indeed, Sanders and Manrodt (2003) found that 83% of a sample of forecasting managers considered "easy understandable results" to be the most important forecasting software feature, while 66% reported dissatisfaction with the software they currently used.

We propose that decision makers will be more likely to accept a DSS when their mental models[1] of the decision environment become aligned with the decision model embedded in the DSS (hereafter referred to as the *DSS model*). The literature provides some support for this view. Gonul et al. (2006) show that confident and long explanations associated with DSS advice can improve user acceptance of that advice. In the context of medical diagnosis of acute cardiac ischemia, Lai et al. (2006) found that a tutorial on the advice given by a clinical DSS increased the use of that advice by emergency care physicians, leading to better patient outcomes. Limayem and DeSanctis (2000) find that system explanations improve group DSS usability, particularly because of improvements in user understanding of decision models.

For mental models to be aligned with the DSS model, decision makers need decisional guidance (Silver 1991). However, Todd and Benbasat (1999) argue that decision makers also have to be induced to exert effort to change decision strategies, which reflect their mental models of the decision environment. We show that a *dual-feedback* DSS, which incorporates feedback *both* about upside potential (i.e., how much more can be gained by internalizing the DSS model) *and* feedback on corrective actions (i.e., guidance on how the manager's mental model should be corrected), would induce more effort from decision makers as well as offer appropriate decision guidance. This combination of effort and guidance then produces significant mental model updating, while single feedback DSS produce little or no updating. Mental model updating, in turn, leads to better subjective DSS evaluations than when little or no mental model updating occurs. While many DSS incorporate some form of feedback, our results show that DSS evaluations only improve after significant mental model updating, which occurs when the DSS incorporates *both* upside potential and corrective feedback.

We proceed as follows. We first present a conceptual framework explaining why the gap between the user's mental model and the DSS model influences

[1] A mental model is an individual's cognitive representation of a domain that supports understanding, reasoning, and prediction (Gentner and Stevens 1983, Norman 1983). The mental model representation of the task is then based on the decision maker's previous experiences and current observations, which provide the framework for how that decision maker performs the task (Wilson and Rutherford 1989, Lim et al. 1997). While the concept of a mental model can consist of many different aspects (including how to use a DSS), we define a decision maker's mental model of a decision domain here narrowly, as a cognitive representation of how multiple decision variables affect performance outcomes.

DSS evaluation. Next we propose a model of how dual feedback on upside potential and corrective actions should influence the updating of users' mental models. We then develop and test specific hypotheses in a realistic, but controlled, experimental setting. We conclude by discussing our research contributions.

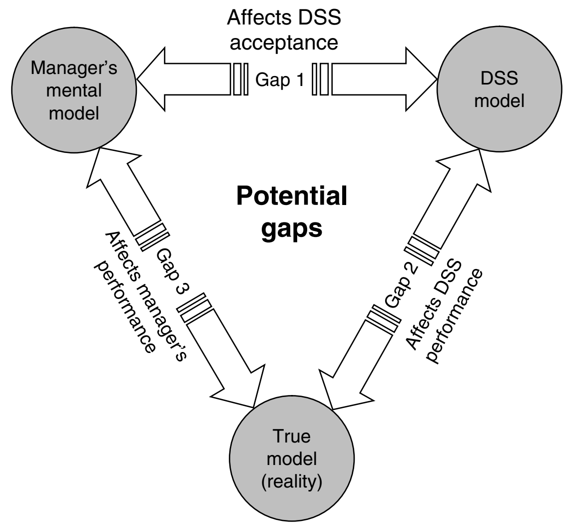
2. Mental Model Changes, DSS Evaluation, and DSS Design

2.1. The Effects of Mental Model Changes on DSS Evaluation

The 3-Gap framework (Figure 1) summarizes our perspective on the DSS evaluation problem.[2] Although we will use managerial decision making as the context, our framework is designed to apply to any data-rich domain where decisions are repetitive and outcomes are uncertain. We hypothesize that the magnitudes of the gaps between three models of the decision environment—the manager's mental model, the DSS model, and the unknown true model (which generates data in the real world, but is only partially observed ex-post)—determine the managers' decisions, the consequent outcomes, and DSS evaluations. To provide high-quality decision support, the gap between the DSS model and the true model must be small[3] (Gap 2 in Figure 1).

When users of a high quality DSS do not understand the rationale behind its recommendations, the gap between the DSS model and the user's mental model of the decision environment is likely to be large (Gap 1 in Figure 1). Consequently, the DSS model's recommended course of action and that implied by the user's mental model are likely to conflict, resulting in decision uncertainty (Einhorn and Hogarth 1980). Based on risk-adjusted preference theory (Keeney and Raiffa 1976), we propose that the objective quality of the DSS is then likely to be discounted by a risk-averse individual to account for the high uncertainty, leading to poorer subjective evaluations. Therefore, one potential source of the DSS evaluation problem lies in the inability of current DSS designs to close the gap between the user's mental model and the DSS model (Gap 1 in Figure 1). As a consequence, we suggest that the greater the change of the mental model in the direction of the DSS model, the better is the evaluation of the DSS that is used to effect the change (formalized later as H1). We focus on how to reduce Gap 1 because we hypothesize that this gap affects the user's evaluation of the DSS. We assume that the DSS model is of high objective quality (small Gap 2) and that it is of better quality than the user's mental model (large Gap 3). (We discuss this assumption in §4.2.)

Figure 1 The 3-Gap Framework: The Effect of Gaps Between Mental Model, DSS Model, and True Model



2.2. Effects of Feedback on Mental Model Changes (Reducing Gap 1 in Figure 1)

We propose that to be recognized by users as valuable, thereby generating favorable evaluations, a DSS must be designed to incorporate characteristics that effect a change in the user's mental model, while improving his/her performance. The change in mental models could be of at least two types—(i) a relatively permanent deep change, or (ii) a transient change that disappears when the DSS is unavailable. We define these changes as follows:

Deep learning is a change in an individual's mental model that endures over time and/or over changes

[2] While gap analysis frameworks have been used in other contexts to understand, diagnose, and improve business and technology performance—see, for example, Parasuraman et al. (1985)—our framework focuses explicitly on the DSS evaluation problem.

[3] We recognize that the real world data generating process is not observable, so any DSS model, however good, is only a stylized representation of the process. The accuracy of a model, thus, might be best judged by how well it predicts the outcomes in a decision environment and/or how well the model fits past data.

in conditions—in other words, a change that concerns "the relatively permanent acquisition of skills, understanding, and knowledge" (Goodman 1998, p. 224).

Shallow learning is a change in an individual's mental model that occurs "only in the presence of external feedback or other conditions of practice, but disappears over time or when the supportive conditions are eliminated" (Goodman 1998, p. 224; also see Kluger and Denisi 1996, p. 278).

Deep learning will tangibly reduce the uncertainty about the DSS recommendations, i.e., reduce (cognitive) dissonance resulting in improved DSS evaluations, whereas shallow learning will not reduce DSS uncertainty. Our main interest in this paper is in how deep learning occurs because we expect it to affect DSS evaluations. Goodman et al. (2004) suggest that deep learning is most likely to occur when individuals are (i) motivated to, and actually exert *effort* to change their mental model, *and* (ii) provided *guidance* on how to modify that mental model, leading to deep learning. We formalize the joint effect of effort and guidance on deep learning as H2. Next we describe how each type of feedback individually influences effort and guidance, and therefore affects learning.

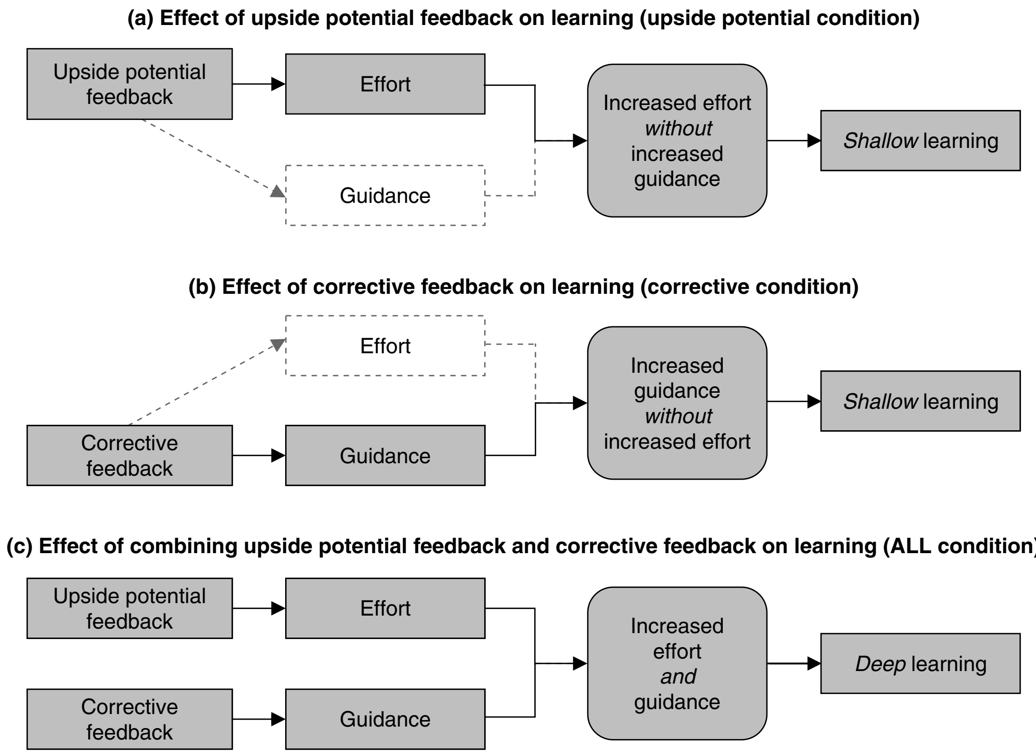
2.2.1. Effects of Upside Potential Feedback on Effort and Learning. Information about upside potential addresses how much better a manager might perform relative to current performance. For example, the sales module of the Siebel CRM system provides a salesperson with information on the sales achieved by the best performing salesperson and the sales levels in the best performing sales territory, providing proxies for upside potential. Chenoweth et al. (2004) show that users of decision support systems exert more effort to learn complex models when they know the upside potential. Upside potential feedback helps managers set specific (and challenging) goals, which drive increased effort to achieve them (Locke et al. 1981, Bandura 1997).

Several researchers have shown that while effort increases with more challenging goals, increased effort does not necessarily lead to deep learning because such goal-oriented behavior can focus the individuals' attention on the self, rather than on the task (Wood et al. 1990). As a result, task-learning processes are not activated (Kluger and Denisi 1996), leading to shallow learning and poor out-of-task performance. Upside potential feedback helps the manager set specific and challenging goals (e.g., match the best salesperson's performance), but does not provide the feedback necessary to learn *how* to perform better. Earley et al. (1990) found that the link between goal-setting, learning, and performance is greatly enhanced when individuals are provided with feedback about how to correct their strategies.

In summary, upside potential feedback will induce increased effort but may direct attention away from task-learning processes, resulting in *increased effort without appropriate learning*. So if upside potential feedback were to be combined with feedback that focuses attention on the task, we would expect managers to exert the increased effort and obtain the guidance necessary to obtain significant deep learning as summarized in Figure 2(a).

2.2.2. Effects of Corrective Feedback on Guidance and Learning. Corrective feedback, also called process feedback (Earley et al. 1990), can improve decision making, particularly in complex tasks, by increasing attention to task-learning processes and improving the quality of decision making (Balzer et al. 1992, Kluger and DeNisi 1996). This attention to task-learning processes improves performance. However, research also suggests that such feedback effects might only be transient—removal of such feedback can bring performance back to where it originally was (Goodman 1998, Goodman et al. 2004), meaning that DSS users will mechanistically implement DSS recommendations when they have a system available, but return to their traditional way of making decisions when the DSS is no longer available. Thus, corrective feedback might only lead to shallow learning because individuals directly adjust behavior by using the feedback rather than using the feedback to understand the task. For example, Atkins et al. (2002) find that if feedback is presented in a way that makes it trivial for decision makers to derive guidelines for action, they won't exert the effort needed to understand the rationale underlying these guidelines. Goodman et al. (2004) note that, "Essentially, feedback does the work for the performers, making it seemingly unnecessary

Figure 2 **Theoretical Framework Relating Feedback to Learning and Evaluation**



Note. Dotted lines indicate expectations of nonsignificant links.

for them to engage in the exploration, information-processing, and recall activities essential for learning." (p. 249).

Thus, corrective feedback directs attention to the task and task-learning process but also leads to less exploration and less effort. The result then is *increased guidance without increased effort*, resulting in low levels of deep learning, as summarized in Figure 2(b).

2.2.3. Effects of Combining Upside Potential Feedback and Corrective Feedback. Our arguments suggest that the two types of feedback should be viewed as *complementary* mechanisms; if the two feedback mechanisms are combined, the result should be an increase in guidance *and* effort, leading to deep learning (i.e., an alignment of the mental model towards the DSS model) as summarized in Figure 2(c).

2.3. Theoretical Model and Hypotheses

Figure 3 summarizes our theoretical framework, relating the two types of feedback to effort and guidance, deep learning, and DSS evaluation. It also provides a summary of the important empirical results discussed later. Our process model is as follows:

$$\mathit{DSS\ Evaluation} = \beta_{01} + \beta_{11} \cdot \mathit{DeepLearning} + \varepsilon_1 \qquad \text{(M1)}$$

$$\mathit{DeepLearning} = \beta_{02} + \beta_{12} \cdot \mathit{Effort} + \beta_{22} \cdot \mathit{Guidance} + \beta_{32} \cdot (\mathit{Effort} \times \mathit{Guidance}) + \varepsilon_2 \qquad \text{(M2)}$$

$$\mathit{Effort} = \beta_{03} + \beta_{13} \cdot \text{UPSIDE POTENTIAL FEEDBACK} + \beta_{23} \cdot \text{CORRECTIVE FEEDBACK} + \varepsilon_3 \qquad \text{(M3)}$$

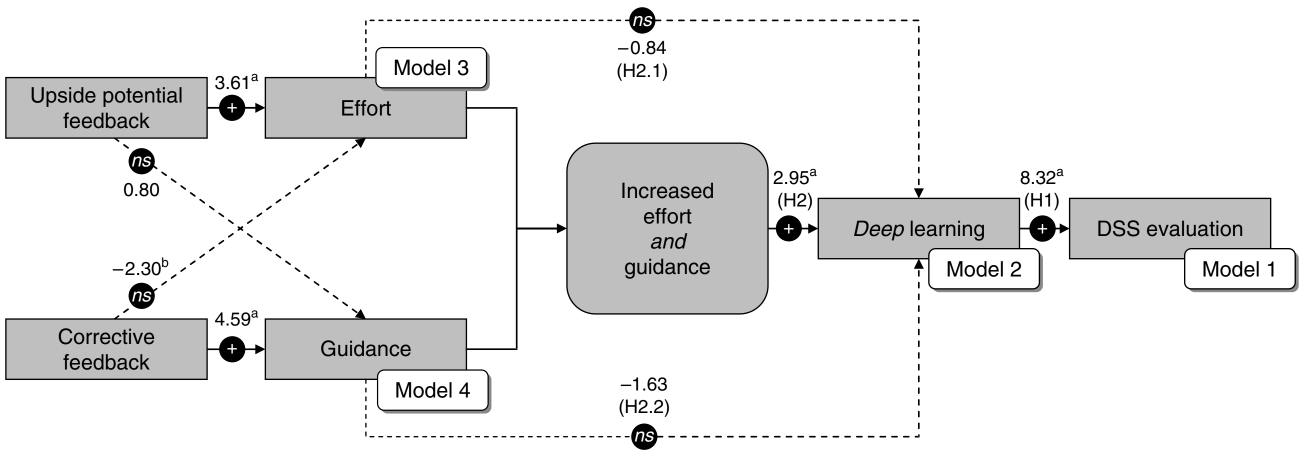
$$\mathit{Guidance} = \beta_{04} + \beta_{14} \cdot \text{UPSIDE POTENTIAL FEEDBACK} + \beta_{24} \cdot \text{CORRECTIVE FEEDBACK} + \varepsilon_4 \qquad \text{(M4)}$$

Our first hypothesis relates DSS evaluation to the relatively permanent change in mental models:

Hypothesis 1 (H1). *An increase in deep learning leads users to provide more favorable evaluations of the DSS. Therefore, we expect* $\beta_{11} > 0$.

We then hypothesize that it is the combination of increased effort and guidance that leads to deep learning.

Figure 3 Connecting DSS Design Characteristics, Deep Learning, and DSS Evaluation (Models M1–M4)



Notes. Dotted lines indicate expectations of nonsignificant links. *ns* stands for *no significant effect expected*; + stands for *positive effect expected*. We report t-statistics and statistical significance ([a]$p < 0.01$, [b]$p < 0.05$, two-tailed), as well as hypothesis numbering when appropriate.

Hypothesis 2 (H2). *The interaction of effort and guidance will have a positive effect on deep learning. Therefore, we expect* $\beta_{32} > 0$.

We also hypothesize that neither effort nor guidance alone leads to deep learning:

Hypothesis 2.1 (H2.1). *An increase in effort without guidance does not lead to deep learning. Therefore, we expect* β_{12} *not to be statistically significant.* (*i.e.*, $\beta_{12} = 0$).

Hypothesis 2.2 (H2.2). *An increase in guidance without effort does not lead to deep learning. Therefore, we expect* β_{22} *not to be statistically significant*, (*i.e.*, $\beta_{22} = 0$).

Per our discussion in §§2.2.1 and 2.2.2, effort increases when a manager is provided with feedback on upside potential, but effort is not expected to increase with corrective feedback, implying $\beta_{13} > 0$, but $\beta_{23} = 0$. On the other hand, guidance is influenced by the presence of corrective feedback but not by feedback on upside potential implying β_{24} to be >0, but $\beta_{14} = 0$. Our expectations about these parameters serve as manipulation checks in our empirical analysis. Models M1–M4 comprise a test of the process model proposed in Figure 3.

3. Empirical Study

We sought to test our hypotheses using a realistic decision environment that is data rich, uncertain, and involves repetitive decisions by managers, criteria met by public charitable organizations soliciting donations via direct marketing (for example, World Vision or National Osteoporosis Foundation). Such organizations typically have access to very large data bases of past donors and prospects, see high uncertainty in response to any specific solicitation and conduct frequent, similar campaigns, leading to decisions that repeat both over time and across prospects.

We asked study participants to assume the role of a direct marketing manager of a large nonprofit charity focused on assisting people affected by natural disasters, and we provided them with a DSS to assist in their donor selection. Their main task was to identify the most attractive donors from a database of past donors for solicitation in a direct marketing campaign. Study participants were MBA students with direct marketing experience, as well as direct marketing managers working at charitable organizations similar to the one in our study.

We designed our empirical study to incorporate a challenging set of criteria. We sought:

(i) a decision environment that would have large data availability, unpredictability, and require repetitive decisions, so that managers would benefit from using a high-quality model-based DSS but not so complex as to be outside the skill range of our research participants;

(ii) a DSS whose underlying model sufficiently captures the real-world phenomenon (i.e., a small Gap 2 in Figure 1);

(iii) a context that would allow us to measure the user's mental model unobtrusively before, during, and after his or her interactions with the DSS;

(iv) a task in which we would be able to embed the DSS with each of the types of feedback (upside potential feedback and corrective feedback), both individually and jointly;

(v) a task that would allow us to measure deep and shallow learning unobtrusively; and

(vi) a task that would allow us to measure the process variables of interest (effort and guidance).

Criteria (i)–(iii) relate to the design of the overall context of the study, whereas criteria (iv)–(vi) relate to the design of the specific experiment to test our hypotheses. Our interest in testing how feedback affects the process of mental model changes (criterion iii) makes the design of a real-world field test challenging. Feedback must be accurate and immediate *across all experimental conditions*. This is difficult to obtain in the real world because of time delays between decisions and results, organizational and environmental noise, and lack of accurate information about what would have happened had other decisions been made (Tversky and Kahneman 1987). Therefore, to obtain both realism and control, we tested our hypotheses under controlled experimental conditions using a frequently occurring and realistic decision problem for which we could offer immediate feedback with known reliability and accuracy.

3.1. Experimental Context

Our experimental context was the solicitation of donations through direct mail for nonprofit or charitable organizations. In the United States alone, direct mail accounts for between \$20 billion and \$25 billion of the charitable educational and social change dollars contributed annually (Lister 2001). Direct marketing managers in charitable organizations typically solicit donations using large databases of potential donors. Each solicitation has a cost attached to it, and donation amount is donor-specific, so that it is critical for the direct marketing manager to identify the most likely (and high value) donors. This situation, in turn, requires the manager to understand the factors that influence the donor's likelihood of donation—that is, a mental model of the drivers of donation. DSS (e.g., MarketMiner Analyst™ website, www.modelingautomation.com) are often used by direct marketing managers to assist them in selecting high potential donors.

3.1.1. Decision Environment. To satisfy criterion (i), we sought a decision environment that would be sufficiently complex, but not outside the skill range of our participants. We designed a direct marketing decision environment complex enough to require the use of a DSS to select customers from a large database (200,000 in our case) of (hypothetical) donors, described on four characteristics—*recency* of donation (the number of quarters since their last donation), *frequency* of donation (the number of donations the donor has made in the past 5 years), *amount* of past donations (the average donation amount, in dollars, observed in the past for this particular donor), and the donor's *age*. The first three characteristics are often used by direct marketing firms in targeting models, typically referred to as Recency-Frequency-Monetary Value (RFM) model. We added age to the model to increase the complexity of the decision environment. Charities commonly use these factors to target donors (e.g., see Schlegelmilch et al. 1997). We modeled the probability, p, that a particular donor would make a donation, if solicited, by a logit function (Agresti 2002) as follows:

$$p = 1/(1+\exp(5-(X/20))), \tag{1}$$

where X is called the donor's "attractiveness" and is given by

$$X = \beta_0 + (\beta_1 \times \mathit{recency}) + (\beta_2 \times \mathit{frequency}) + (\beta_3 \times \mathit{amount}) + (\beta_4 \times \mathit{age}). \tag{2}$$

The parameters of the "true" data generating model were $\beta = \{20, -20, 40, 10, 30\}$. We informed participants that donors were more likely to donate if they (1) had donated more recently, (2) had donated more frequently in the past five years, (3) had donated greater amounts, and (4) were older.

We generated a database of customers to satisfy two criteria. (1) The probabilities of donation in our database should be similar to those observed in actual not-for-profit databases. (2) The characteristics should be generated such that each donor could be described on a 0-to-100 attractiveness scale with an average at the midpoint. The latter criterion ensured that we could subsequently ask participants to rate each donor on the same scale. To satisfy these two criteria, we generated donor characteristics from uniform distributions between 0 and 1 (after rescaling to account for differences in measurement units) independent of one another and incorporated these values within the functional form of the logit function described in Equations (1) and (2). As a result, a donor's true attractiveness varied between 0 and 100 with an average of 50, and donors' probability of donation varied between 0.67% and 50% with an average of 7.6%. Although average response rates, in practice, vary widely for different charities, an average response rate of 7.6% falls within industry averages for "warm" donors (see www.fundraising.co.uk/forum/thread.php?id=500).

Per criterion (ii), we sought a DSS model that would be close to the true data-generating model. Therefore, we designed Gap 2 to be small by constructing a DSS model that was identical to the true model in terms of weight of each factor. However, to ensure that actual donations could be predicted only approximately by the DSS model, we added a random noise term to the true model in Equation (1).

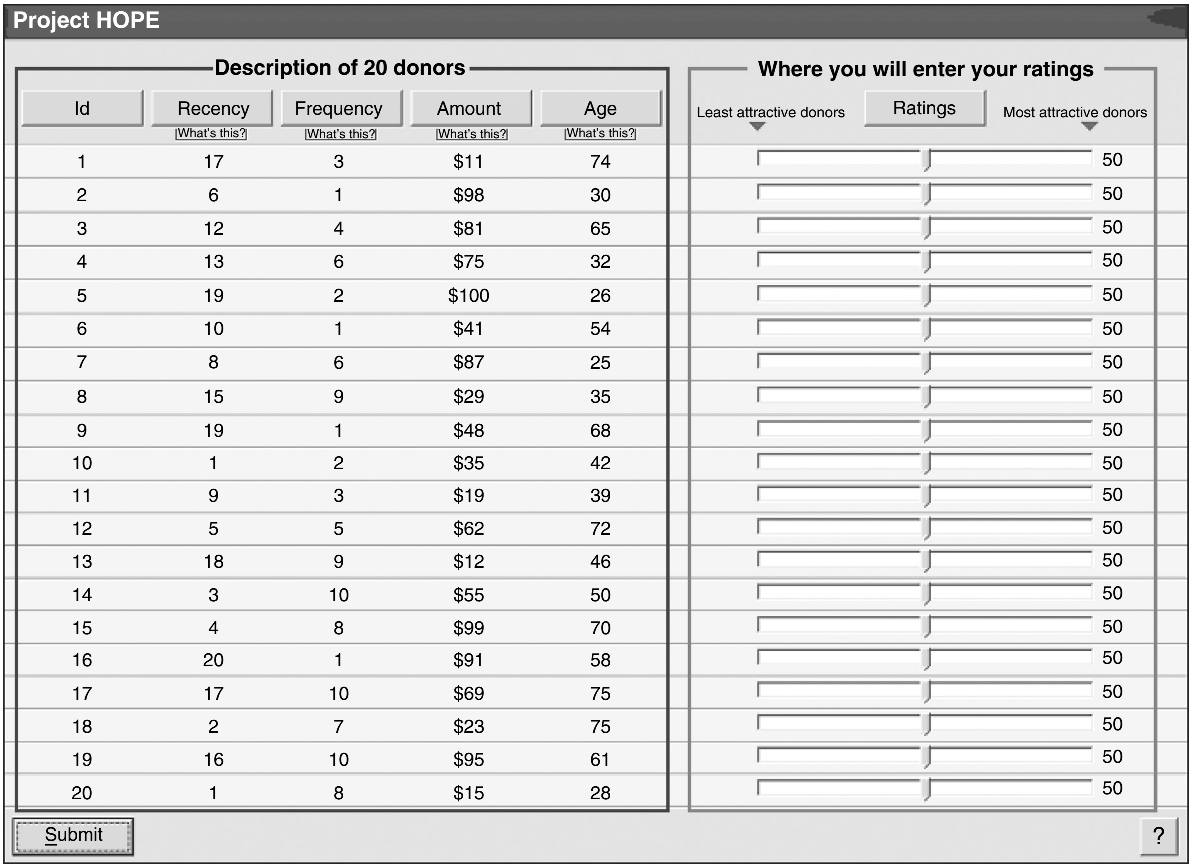
3.1.2. Calibrating Participants' Mental Models. To satisfy criterion (iii), we devised an unobtrusive and unbiased mechanism to calibrate each participant's mental model. We designed our study so that each participant made a sufficient number of decisions at each stage of study. This requirement allowed us to unobtrusively calibrate the participant's mental model, similar to Kunreuther's (1969) work on estimating managerial decision coefficients. Our unobtrusive approach minimizes potential biases compared to directly asking participants to reflect on their mental processes (Norman 1983). We asked each participant to rate 20 donors from the database on a 0 to 100 scale reflecting how attractive each donor was for selection in a marketing campaign. These 20 donors were described along the four drivers of donation behavior—recency, frequency, donation amount, and age (see Figure 4 for a screen shot of the task). This rating process (after rescaling and sorting) corresponds to the typical scoring mechanism that emerges in most direct marketing DSS (see D. Shephard Associates 1999).

To measure the participant's mental model, we statistically related their donor ratings to the descriptions of the 20 donors, thus inferring the implicit weights participants placed on the four factors.[4] Once a participant submitted his or her ratings, we estimated a linear regression model to determine the implicit weights ($\beta'_0, \beta'_1, \ldots, \beta'_4$) that participant placed on recency, frequency, donation amount, and age. We then applied this calibrated mental model to the larger database of 200,000 donors to determine who to solicit. We told participants that each solicitation costs \$2 and, if successful, would generate a constant \$20 donation (to keep the task within participant skill range, criterion i), yielding a profitability threshold of 10% probability of donation. We applied the estimated mental model to the entire database, computing X' and p' for each of the 200,000 donors, and soliciting those donors with $p' > 0.1$. In addition to the marginal costs of solicitation, the fundraising campaign was subject to fixed costs of \$10,000. To determine whether a solicited donor actually makes a donation, we draw a random number z from a uniform distribution $[0, 1]$ for each donor, and each solicited donor makes a donation of \$20 if z is less than the true probability of donation (p). Note that if participants provided perfect scores ($X' = X$), mental model parameters would be equal to true parameters ($\beta' = \beta$), and the solicitation strategy would be optimal.

To assist participants in their decision making, we provided them with a DSS to select attractive donors from a database. We addressed the issue of incentive alignment by informing participants in all conditions that the amount of money they earned would be directly proportional to their financial performance. Participants were paid 0.015% of their financial performance on Task 1 and Task 2 (described more fully

[4] To estimate this relationship, there must be sufficient variation in the description of the 20 donors on each of the four factors and the factors must not be multicollinear, allowing for independent estimation of each weight. While a fractional factorial design is typically used in such cases, the number of profiles that our participants would have to rate made that approach infeasible. Therefore, we generated donors' characteristics (recency, frequency, etc.) so that extreme values were represented more often in the sample than in the population, while spanning the entire parameter space. To avoid multicollinearity, we randomly permuted donors' characteristics in the participant's rating sample, until no intercharacteristic correlation was higher than 0.15.

Figure 4 DSS Interface, Illustrating the Respondent Task



Id	Recency	Frequency	Amount	Age	Ratings
1	17	3	$11	74	50
2	6	1	$98	30	50
3	12	4	$81	65	50
4	13	6	$75	32	50
5	19	2	$100	26	50
6	10	1	$41	54	50
7	8	6	$87	25	50
8	15	9	$29	35	50
9	19	1	$48	68	50
10	1	2	$35	42	50
11	9	3	$19	39	50
12	5	5	$62	72	50
13	18	9	$12	46	50
14	3	10	$55	50	50
15	4	8	$99	70	50
16	20	1	$91	58	50
17	17	10	$69	75	50
18	2	7	$23	75	50
19	16	10	$95	61	50
20	1	8	$15	28	50

in the next section), in addition to $15 for participating in the study. The performance-based incentive was identical across all conditions.

3.2. Design of the Experiment (Manipulations and Measurements)

We summarize the sequence of steps in our experiment in Box A of Figure 5. Our experiment consisted of three main parts, addressing study design criteria (iv), (v), and (vi), respectively.

3.2.1. Part 1: Using the DSS. In Part 1 of the study, we asked participants to rate the same 20 donors in each of ten simulations. The participants had access to a DSS to help them determine the best possible ratings. In each simulation, participants rated the 20 donors, submitted those ratings to the DSS simulator, and obtained the DSS prediction of the performance of the campaign based on the participants' donor ratings. In the background, we calibrated the regression relationship between the participants' ratings and the description of the 20 donors on the 4 factors. The DSS was therefore both a support tool for users to make decisions and also a research tool to measure users' mental models.

We varied the feedback provided by the DSS to reflect the two types of feedback under study. We varied upside potential feedback at two levels (present or absent) and corrective feedback at two levels (present or absent), for a design with four cells. Both types of feedback were absent in the control condition. Our four conditions were:

1. "CONTROL CONDITION": The participant was only informed of the expected performance of the donor ratings. For example:

> The DSS predicts that a marketing campaign based on your ratings would generate $76,654 in revenue.