

# **Diagnosing Decision Quality**

**Michael J. Davern**

Department of Accounting & Business Information Systems  
The University of Melbourne, Australia

**Ravi Mantena**

William E. Simon Graduate School of Business Administration  
University of Rochester, Rochester, NY.

**Edward A. Stohr**

Howe School of Technology Management  
Stevens Institute of Technology, Hoboken, NJ.

Original Version: March 2007 (**DSS#07-04-1932**)

Revised: September 2007

## **ABSTRACT**

Human decision making is error-prone and often subject to biases. Important information cues are misweighted and feedback delays hamper learning. Experimentally, task information has been shown to be valuable in improving decision making. However, such information is rarely available. Generalizing from lab-based approaches, we present a new field methodology for the “Diagnosis of Decision Quality” (DDQ) that helps decision makers discover such information. We illustrate our approach in the context of hotel revenue management and demonstrate how it can identify context-specific systematic errors in decision making in a manner that facilitates adaptive changes and improved performance.

**Keyphrases:** Decision quality, Decision support, Human judgment, Cognitive feedback, revenue management.

## 1. INTRODUCTION

Assessing the quality of decisions after the fact is a pervasive human preoccupation. However, systematic studies of how the wisdom of prior decisions can be assessed are relatively rare [21]. In general, there is a tendency to judge decisions based on their *outcomes* – if the outcome is good then, ipso facto, the decision was good. An alternative view, and the one that we will adopt in this paper, is that decisions should be judged good or bad according to the quality of the process by which they were made [21].

The substantial literature in judgment analysis [9, 34] has ably demonstrated that improving human decision making processes is a non-trivial exercise. Humans, while adaptive, often ignore relevant information or weight it inappropriately [15, 7]. Providing useful decision support tools is also not a sure-win solution, as users often erroneously discount the value of the quantitative models they embody [11, 12, 13, 14, 16]. To improve decision performance, *diagnostic information* and *feedback* should be provided to decision makers – information that can assist them in revising their decision processes rather than simply reporting their performance. In laboratory settings, the provision of task information, such as cue-criterion relationships, has been shown to be key to improving performance in judgment tasks [2]. In real organizational settings however, such information is difficult to obtain for all but trivial tasks.

In this paper, we present a new methodology for the “Diagnosis of Decision Quality” (DDQ) in complex real world decision contexts. The objective is two-fold. First, to identify systematic strengths and weaknesses in a decision process, and second, to improve decision performance by facilitating the discovery of outcome-pertinent task information. We base this method on an analysis of the generic diagnostic characteristics of the well known lens model [6], which is the basis of decades of laboratory-based judgment analysis studies (good reviews are available in [1], [9]). We show how these diagnostic characteristics can be mimicked in real world settings using our methodology. We illustrate the application of our approach in the context of hotel revenue management using daily data obtained from an international upper-upscale hotel chain.

The paper is structured as follows. We start with an analysis of the defining characteristics of laboratory-based judgment analysis methods; this serves as a guide for the development of new judgment

analysis approaches (both field-based, such as ours, and lab-based). Next, we present our methodology and provide a proof of concept [19, 23] in its application to revenue management at a major hotel chain. We also demonstrate the practical value of our methodology by identifying context-specific systematic biases in decision making in the hotel chain that suggest the generation of adaptive changes (e.g., decision support interventions) to improve subsequent performance. Finally, we discuss the contributions of our methodology and point to some applications in areas other than revenue management.

## **2. THEORETICAL BACKGROUND AND METHODOLOGY DEVELOPMENT**

### **2.1 The Role of Feedback in Improving Performance: Experiments in Human Judgment**

**Human judgment and the lens model.** Human judgments – opinions about the current or future value of some aspect of the world – are central to much of human decision making. As Yates [30, p.6] notes: *“Judgments form the cornerstone of most decisions. The quality of those decisions can be no better than the quality of the judgments supporting them.”* Consequently, we draw on studies of human judgment and learning for the theoretical development of our approach for diagnosing and improving decision quality.

Experimental studies of human judgment in multiple cue probabilistic environments typically adopt either a Bayesian or a regression-based approach for modeling decision making [27]. Originating in the work of Brunswik [6], the lens model, which employs regression, has been used extensively in laboratory studies of human judgment for fifty years. The fundamental philosophy behind this approach is that a full understanding of judgment performance requires modeling not just human judgment, but also the task environment in which the judgments are made, and the relationship between the two. Many of the lens model studies have been explicitly aimed at understanding the effects of feedback on performance. We build on this tradition and frame our theoretical development in terms of the lens model. Figure 1 shows a common formulation (e.g. [8]) of the lens model.

-----  
Insert Figure 1 Here  
-----

Under the lens model, the environment is modeled as a linear combination of a set of information cues ( $x_i$ ), which may be correlated ( $r_{ij}$ ) and that are predictive of a criterion variable ( $Y_e$ ) (e.g. temperature, a cue, might be predictive of the criterion variable sales of ice cream). The correlation ( $r_{ei}$ ) between a cue and the criterion is referred to as the ecological or predictive validity of the cue. Because the environment is complex and stochastic, this linear model of the environment is not perfectly predictive of the criterion. The extent of unpredictability in the criterion is indicated by the correlation ( $R_e$ ) between the actual outcome ( $Y_e$ ) and the predicted outcome ( $\hat{Y}_e$ ) ( $R_e=1$  implies a certain and perfectly linear environment).

In laboratory studies, the decision maker (experimental subject) responds to multiple sets of values for cues and, in each case, makes a prediction ( $Y_s$ ) of the criterion. Using regression, a linear model for each subject is constructed to find the weight ( $r_{si}$ ), on average, implicitly applied to each cue by the subject in predicting the criterion;  $r_{si}$  is referred to as the utilization validity of the cue. Because subjects are often inconsistent, this linear behavioral model is not perfectly predictive of their judgments. Consistency is indicated by the correlation ( $R_s$ ) between the subject's prediction of the criterion ( $Y_s$ ) and that suggested by the model describing the subject's behavior ( $\hat{Y}_s$ ) ( $R_s=1$  implies perfect consistency).

In the lens model, performance is measured by the correlation ( $r_a$ ) between the criterion ( $Y_e$ ) and the subject's prediction ( $Y_s$ ). A subject's knowledge is measured by  $G$ , the correlation between the prediction ( $\hat{Y}_e$ ) of the environment model and the prediction ( $\hat{Y}_s$ ) of the behavioral model [18].  $G$  is essentially a measure of how well a subject's cue utilizations correspond to the predictive validity of the cues – how appropriate is the subject's weighting of the cues. Judgment performance is given by the lens model equation:  $r_a=GR_sR_e$  (assuming uncorrelated residuals [29]) and is a function of the subject's knowledge ( $G$ ), his consistency in applying his knowledge ( $R_s$ ) and environmental uncertainty ( $1-R_e$ ).

**Types of Feedback.** Simple outcome feedback (e.g.,  $r_a$  in the lens model formulation), while the most readily available information both in the lab and in the field, is also the least valuable for learning. Outcome feedback has been found to assist learning only in very simple situations with two or fewer cues linearly related to a criterion [9]. Indeed, in a review of several studies, Brehmer [4] presents evidence

that simple outcome feedback can actually be harmful to learning. This is not surprising since outcome feedback is non-diagnostic, telling the decision maker only the quality of the results, but nothing about how to adapt decision processes to improve performance. More formally, this is evident from the lens model equation. Given unfavorable performance (low  $r_a$ ), it is impossible to identify its cause from among the competing alternatives -- the decision maker's knowledge ( $G$ ), inconsistency in applying the knowledge ( $1-R_s$ ), or the inherent unpredictability of the criterion ( $1-R_e$ ).

An alternative to simple outcome feedback is Cognitive Feedback. Balzer, et al [1], (following [17], [28]) describe three distinct components of cognitive feedback within the lens model structure; Task Information (TI), Cognitive Information (CI), and Functional Validity Information (FVI) (see Table 1).

-----  
Insert Table 1 Here  
-----

In a comprehensive review of studies employing the lens model, Balzer, et al [1] conclude that TI (e.g., cue-criterion relations) appears to be the most effective in improving performance. In a later study, Balzer, et al [2] empirically compared five different feedback conditions (TI only, CI only, TI+CI, TI+CI+FVI, and no feedback). TI was again found to lead to significant improvements in performance, CI alone never outperformed no feedback, and the addition of CI or FVI on top of TI did not lead to additional performance gains.

Experimentally, therefore, TI is the most useful type of feedback for learning. This is not surprising because it provides the most direct guidance to decision makers in adapting their existing decision processes. Specifically, the provision of information about cue-criterion relationships tells decision makers exactly how much they should weight different cues in their judgments (in this regard it can be seen as cognitive feedback, because it provides guidance as to a normative ideal for cognition – cue utilizations that correspond to ecological validities). In practice however, TI is difficult to obtain. Indeed, if the cue-criterion and inter-cue relationships are well known, or easily obtainable, then the judgment task becomes quite trivial and can be readily automated.

What real world decision makers require then, is diagnostic feedback that helps them discover TI – the relevant cue-criterion relationships. To be truly diagnostic, the feedback should be focused on aiding discovery of task information for the cues that are misweighted or misjudged. Diagnostic decision support tools thus need to be both focused and economically appropriate – reflecting the context specific weaknesses in decision making processes and the need to balance the costs of data acquisition and analysis against the benefits of improved decision making. The *diagnosis of decision quality* (DDQ) methodology described below is designed with these objectives.

## **2.2. From the Lab to Practice: Characteristics of a Methodology for Diagnosing Decision Quality**

**Theoretical considerations.** There are several key characteristics of the lens model that underlie its usefulness as a framework for analyzing judgments and building a practical methodology for diagnosing decision quality. Towards this end, it focuses attention on the following aspects:

- (i) A criterion ( $Y_e$ ) against which performance may be assessed ( $r_a$ )
- (ii) Separation of random variation in performance from systematic deficiencies (i.e.,  $1-R_e$  vs poor  $G$ )
- (iii) Identification of relevant cues in the environment ( $X_i$ )
- (iv) Knowledge of the “true” weights associated with the cues (the predictive validities  $r_{ei}$ ) relative to the weights assigned by the decision maker (the cue utilizations  $r_{si}$ ).

Item (i) is essential for identifying decision *outcomes* as favorable or unfavorable. However, this in itself does not aid in improving performance as outcomes are a joint product of the quality of the decision maker’s judgments (indirectly, their knowledge), and the inherent uncertainty in the task environment. The classification of *judgments* (rather than outcomes) as good or bad therefore requires controlling for the effect of the environment, which can be accomplished using item (ii). Items (iii), and (iv) enable an analysis of the decision maker’s knowledge, and are useful in diagnosing systematic strengths and weaknesses in the decision process, thereby finding opportunities for improvement.

Thus to *evaluate* performance, a method for the diagnosis of decision quality must specify an ideal benchmark (criterion) for performance and a means for separating out systematic deficiencies in

judgment from random variations in performance outcomes. To *diagnose* performance (and consequently provide feedback useful for improving performance), such a method must also help decision makers identify relevant cues ( $X_i$ ) that are misweighted, and that are driving unfavorable performance outcomes.

**Pragmatic Constraints.** The regression-based approach of the lens model has provided a great deal of insight into human judgment in laboratory settings. However, as we move from the lab to the real world, the objective is different – to assess and improve decision performance in specific decision contexts, rather than to broadly understand human judgment. There are also inherent difficulties in achieving each of these steps in the field (e.g., simply obtaining an ideal performance benchmark can be non-trivial). The complex and highly non-linear nature of real-world data further limits the usefulness of directly applying the lens model. In addition, other pragmatic constraints must be taken into account when developing an effective field methodology for the diagnosis of decision quality. Specifically:

- (i) Multiplicity of objectives: Any useful methodology must recognize the multiple competing objectives that are typical of real world decision situations and consequently decision makers' performance needs to be evaluated in terms of multiple measures.
- (ii) Data availability and/or acquisition costs: A good methodology should allow for varying depths of analysis, conformant with the availability of data and its cost of acquisition.
- (iii) Understandability of methodology and results: The feedback provided by a methodology must be plainly understandable by the decision makers themselves. Similarly, the methodology itself must have face validity and thus not be overly complex if it is to be accepted by decision makers and lead them to revise their decision processes accordingly.

In the following, we use the lens model as a guiding framework to develop a field methodology that explicitly incorporates the desirable characteristics identified in this section, while freeing itself from the limitations imposed by a regression-based approach.

### 2.3 An Overview of the Diagnosis of Decision Quality (DDQ) Methodology

At a generic level, the DDQ methodology can be described in two phases of two steps each (see Table 2). This section provides a brief overview of the structure of the methodology and then in Section 3 we provide a more detailed description and illustrate its application in the hotel revenue management context.

-----  
Insert Table 2 Here  
-----

**The evaluation phase.** The purpose of the evaluation phase is to assess decision performance. Performance here refers to the quality of the decisions, rather than the quality of the outcomes. In practical decision contexts, the outcome is dependent on both the environment (exogenous factors) as well as the decision maker's choices (endogenous factors). Therefore the quality of the outcomes is at best stochastically related to the quality of the decisions.

The first step in the analysis is the identification of relevant performance measures (criteria -  $Y_e$ ) and benchmarks for measuring performance ( $r_a$ ). Since in practice, performance is multi-attributed (due to multiple objectives), a portfolio of measures will likely be required. For each of these measures, a benchmark or criterion level needs to be defined. In doing so, one should carefully consider the exogenous and endogenous factors driving performance and ensure that the benchmarks chosen reflect the quality of the decision process rather than factors outside the control of the decision maker (G vs.  $R_e$ ). Achieving this separation is in general quite hard. However, in our context, this is facilitated by two factors. First, our objective is not to arrive at an absolute measure of the decision maker's performance, but rather, to identify systematic deficiencies that can be targeted for improvement. Second, our methodology is based on a post-hoc analysis of data, which means that both the values of the exogenous variables and the endogenous decisions made are known at the time of the analysis. This allows us to identify (ex-post) the *optimal decisions that should have been made*, given the values of the exogenous variables, and then compare these to the *actual decisions* to identify opportunities for improvement. This comparison automatically controls for the effect of exogenous variables as both the actual and the optimal

decisions are subject to the same set of values on these variables. The fraction of decisions that were optimally made provides a good measure of the *decision quality*. Alternately, the fraction of decisions that were suboptimal is a measure of the opportunity for improvement. Also even though multiple performance measures may be identified, they need not all be calculated/estimated, nor benchmarks defined for them at this stage, unless diagnostically necessary. This is a critical strength of our approach.

The second step in the evaluation phase is to develop a classification tree (the DDQ Tree) with branches corresponding to the selected performance measures and benchmark values. At the simplest level, the tree classifies the outcome from each decision instance as either *favorable* or *unfavorable* to the outcome desired by management. More generally, the tree provides a set of decision outcome classifications that separate performance fluctuations due to exogenous factors from systematic effects (strengths/weaknesses in the underlying decision making processes or task difficulty).

**The diagnosis phase.** The objective of the diagnosis phase is to identify systematic strengths and deficiencies in the decision making process (functional validity information) as well as to identify appropriate task information (TI). For the task information to be most effective at improving decision performance, the information should relate to cues that are inappropriately weighted by the decision maker. Systematic deficiencies in decision making can be deduced from the distribution patterns of the DDQ tree classifications. This is the first step in the diagnosis phase. For example, at the simplest level with a purely random decision process, we would expect an equal split between favorable and unfavorable classifications. Any significant deviation from this reflects some systematicity (favorable or unfavorable) in the outcomes, and consequently reflects a systematic discontinuity between the decision maker's behavior and the task environment. Of course this analysis is much more effective with finer grained classifications, as we will illustrate in our application of the approach in section 3 below.

A more elaborate analysis aimed at discovering the drivers of the decision outcome patterns can be obtained by analyzing the decision process instances that fall within each DDQ classification. This is done in step two of the diagnosis phase, which requires two things - the identification of relevant cues and DDQ control chart analysis. Active involvement from decision makers in the domain is needed both

to identify relevant cues and to interpret the charts to discover systematic deficiencies. The cues in this analysis need not be the same as those consciously used by the decision makers. Neither do they need to be direct causes of the systematic patterns in DDQ classifications. They merely need to be related to the cause sufficiently so that the DDQ control chart can facilitate the discovery process by the decision maker – a process that may itself lead to further data analysis, including the identification of additional cues for the plotting of DDQ charts. Ultimately, in lens model terms, this iterative process leads the decision maker to discover task information in the form of cues ( $X_i$ ) for which the cue utilizations ( $r_{si}$ ) deviate most significantly from the predictive validities ( $r_{ei}$ ) (i.e., the cue is mis-weighted).

### **3. THE DDQ METHODOLOGY APPLIED: A PROOF OF CONCEPT**

#### **3.1 The Application Context: Revenue Management at Alpha Hotels**

Hotel revenue management provides an excellent context for illustrating the application of our DDQ methodology. Revenue management is a recurring task of substantial economic significance. While sophisticated computer-based revenue management systems have become increasingly prevalent, the end decision still involves a great deal of human judgment [11, 12]. Revenue management is typically applied in industries with a fixed short-term capacity, high fixed costs, low variable costs, and where the opportunity for revenue is perishable. Examples include airlines, hotels, car rentals and advertising. The basic idea is to capture as big a fraction of the revenue opportunity as possible through effective price discrimination [10]. Customers are divided into segments based on their willingness to pay. The fixed short term capacity is then allocated to these price segments optimally with the twin objectives of (i) maximizing the fraction of sales to the higher priced segments, and (ii) minimizing the extent of unused capacity. Thus, with ample unfilled capacity, a firm practicing revenue management will increase its low price segment sales. Conversely, when capacity is scarce, it may be better off turning away low price customers in the expectation that a sufficient number of high price customers will be forthcoming before the product perishes and the revenue opportunity is lost. Revenue management therefore involves two critical tasks, a segment-wise demand forecast and an optimal capacity allocation given that forecast.

Hotel revenue management is a complex task, involving repeated decisions over time (and consequently opportunities for learning) with a sizeable number of relevant cues and a large feedback delay between actions and decisions because of the extended time period (the “booking cycle”) over which reservations arrive. Three classes of actors are involved in hotel revenue management: hotel senior management, revenue managers (RMs), and booking agents. In general, senior management sets the longer term strategic policies, the RMs manage parameters that govern the room offerings available for any given future date (“room night”) and the booking agents conduct reservation transactions with potential customers. Here, we are concerned only with the decisions made by the RMs.

Following revenue management practices, hotels offer rooms at a number of price points or buckets, for example, premium, corporate, contract, and super-saver. When a target “room night” is sufficiently far in the future, all price buckets will normally be open. As a target room night approaches and the accepted reservations approach capacity, the RM may close the cheaper buckets in an attempt to sell the remaining rooms to only the highest paying customers. Essentially the RM seeks to fill the hotel with the highest paying customers possible, while also trying to ensure that no rooms remain unsold.

Revenue management price buckets do not correspond in a precise fashion to physical rooms. For example, a given room may be charged at a rate corresponding to either the “Premium” or “Super-Saver” bucket, depending on the time it is booked relative to the target date. In effect, this means that the sizes of the buckets and/or their availability can vary over time. The actual price of a given bucket is usually fixed for a given day of the week for a period of time (e.g., a super-saver rate of \$120 on all Mondays during peak season). RMs’ day-to-day decision-making thus involves opening and closing priced buckets given a fixed rate structure. For example, an aggressive or optimistic RM might close the “Super-Saver” bucket earlier in the reservation cycle for a given room night than a more conservative RM.

Hotels routinely track at least three different performance measures: *Occupancy*, *Average Daily Rate (ADR)*, and *Revenue per Available Room (RevPAR)*. To understand these measures, consider a hotel with 200 available rooms. Assume that on a given day it has managed to fill 170 of the rooms, not all at the same rate. Further assume that the total room revenue it receives for that day is \$22,100.

- *Occupancy*: Occupancy, usually defined as the percentage of available rooms that are occupied on a given room night ( $= 170/200$  or 85% for the above example), is a traditional measure of hotel performance. Improvements in occupancy are indicative of good revenue management decision making to the extent they are not driven by exogenous increases in market demand or obtained by unnecessarily lowering rates, thereby unduly sacrificing revenues.
- *Average Daily Rate*: ADR, calculated as total room revenue divided by number of rooms occupied ( $= 22100/170 = \$130$  in the above example), is another stalwart of hotel performance measurement. For ADR to be a useful measure of the quality of revenue management decisions, it must not have been inflated by sacrificing occupancy or by changes in market conditions.
- *Revenue Per Available Room (RevPAR)*: RevPAR ( $= \text{Occupancy} \times \text{ADR}$ ), combines ADR and Occupancy to capture the effects of the trade-off between the two measures. In the hotel industry RevPAR is widely considered the critical performance measure. For the above example,  $\text{RevPAR} = 22100/200 = \$110.50$  ( $= 130 \times 85\%$ ).

The above measures of RM performance, by themselves, are primarily evaluative rather than diagnostic. Even for outcome evaluation, using these measures alone makes it difficult to connect performance to decision processes as opposed to exogenous market-related effects. However, they serve as useful starting points for our methodology.

***The research site and data.*** We illustrate the application of the DDQ methodology using daily data obtained from an international upper-upscale hotel chain that we will refer to as Alpha Hotels. Each hotel typically comprises several hundred guest rooms. Organizationally, Alpha is heavily decentralized with a mix of franchised, managed and chain-owned facilities. Individual hotel managers have substantial autonomy, and consequently, a key problem faced by the chain head office is to convince the hotels of the need to spend significant funds on revenue management decision support.

Alpha provided us daily data for up to 2 years for all the hotels in the chain on variables such as occupancy by rate class, average daily rate, denials (discussed later), and cancellations. Data was provided in the form of Excel spreadsheets submitted to the chain head office by the individual hotels. Reflecting the costs of data collection, several of the hotels did not track complete data. Our analysis focuses on the two hotels that the chain identified as their leading performers: “Big City” located in a major US metropolitan area and “Southern Country” located in a southern US regional center.

### **3.2 Applying the Methodology**

To start with, we need to define an outcome and identify the set of decisions that lead to the outcome. Here, an outcome is defined as the revenue realized for a reservation night, and the RM decisions that affect this outcome are the opening and closing of price buckets. While we treat the opening/closing of price buckets for a reservation night as if it were a single independent decision, in reality, it involves multiple decisions over a number of days leading up to the reservation night. Also, the decisions for different reservation nights are not completely independent due to multiple-night stays. However, the methodology is quite robust to both these situations. In the former situation, the DDQ methodology simply provides diagnostic feedback about the decisions as a set (diagnostic analysis is thus only hampered when any systematic biases play out inconsistently over the set of decisions). In the latter situation, if interdependencies are significant, it may be necessary to evaluate outcomes contingent on past history. In our case, the revenue realized on a given reservation night is treated as independent since multiple night stays are not significant in our data set.

#### **The evaluation phase: Step 1 – Performance measures and benchmarks**

The first step in the evaluation phase of our methodology (see Table 2) involves identifying performance measures and establishing benchmarks. We described three possible measures of hotel revenue management performance in the previous section. Of these, while RevPAR is the most comprehensive measure, using ADR and Occupancy separately affords greater granularity in connecting outcomes with decision processes and provides greater diagnosticity. Therefore we choose these two measures.

Establishing meaningful performance benchmarks for these measures is less straightforward, and requires a more detailed description of hotel revenue management decision making. Assuming that the revenue manager has perfect information about demand, the task can loosely be defined as:

**Objective:** Maximize Revenue

**Constraints:** Hotel Capacity, Market Demand

**Action:** Pricing Strategy Selection: Specifically opening and closing price buckets

**Optimal Decision Rules:**

IF Unconstrained Demand<sup>#</sup>  $\leq$  Capacity THEN do nothing

IF Unconstrained Demand  $>$  Capacity THEN close out price buckets where price is below  $P^*$

WHERE  $P^*$  is defined such that (Total Demand at  $P^* =$  Capacity).

<sup>#</sup> Unconstrained Demand is the number of rooms that could be sold if the hotel had infinite capacity i.e., when no customer is denied a room irrespective of her willingness to pay.

Of course, in practice, these simple rules are difficult to instantiate primarily because RMs do not have perfect information about the market demand curve they are facing and therefore cannot determine unconstrained demand and solve for  $P^*$ . Hence, RMs must make *judgments* about expected demand. Applying the rules is also complicated in practice because the hotel's pricing structure is not continuous – the hotel may not have a price bucket with the exact price  $P^*$  that would equalize supply and demand.

Two key benchmarks in revenue management are full occupancy and the Market Demand Curve. The market demand curve for a room night describes the numbers of customers of different willingness to pay that are likely to seek a room. This enables one to calculate an exogenous upper bound on the revenue an RM can realize for the hotel. Therefore, given our focus on understanding the effectiveness of the decision process, rather than the absolute quality of the event outcomes, any evaluation has to factor in the market demand curve when classifying decision outcomes as good or bad.

While it is not possible to directly reconstruct the actual Market Demand Curve for a particular reservation night, even ex post, it is possible to estimate unconstrained demand (an exogenous variable), and in fact, most hotels routinely track the necessary data. Unconstrained demand is defined as occupancy plus denials. Denials are prospective customers who are turned away in the reservation

process due to the lack of availability of a room at a price bucket they are willing to accept. Although the hotel may not be full, these low paying customers are denied a room in anticipation of higher paying customers arriving later in the booking cycle. Given unconstrained demand, the market demand curve is obtained by splitting the unconstrained demand based on the distribution of willingness to pay. This distribution can in turn be estimated from data on days for which there are no denials, when the sales mix of the hotel is the same as the market mix. This estimation assumes that the distribution is stationary for a given day of the week within a given season and independent of the total unconstrained demand.

Denials, ADR, market demand and Occupancy data are adequate for commencing the second step of the evaluation phase – the construction of the DDQ tree.

-----  
Insert Figure 2 Here  
-----

### **The evaluation phase: Step 2 – The DDQ tree**

A DDQ classification tree is constructed using the measures, benchmarks and exogenous variables identified in Step 1. The tree is then used to classify each outcome into diagnostic classes.

Before constructing the actual tree, we first exclude outcome instances in which the quality of the decision process is not reflected in the outcomes. In the hotel revenue management context, if there are no denials ( $\text{Denials} = 0$ ) for a given reservation night, then whatever decision the RM made about closing price buckets did not have any impact on the final outcome. This is because no customer was turned away due to the unavailability of a room at the price she was willing to pay – an outcome which would have resulted even if there was no RM and consequently no closing of price buckets. We have no basis on which to judge the quality of the RM's decisions in these cases, at least not with the data we have available (e.g., assessment would be possible if detailed records of demand forecasts exist), and therefore these outcomes are excluded from the initial set before proceeding to the actual DDQ tree.

Figure 2 shows the DDQ tree for the hotel revenue management context. In generic form, the diagnostic classes in a DDQ tree are as follows:

- *Favorable*: The performance outcomes were favorable relative to a benchmark i.e. they resulted in improved outcomes for the hotel.
- *Unfavorable*: The performance outcomes were unfavorable relative to a benchmark i.e. the outcomes were sub-optimal and there is room for improvement. This can be further broken down into:
  - *Optimistic or Unlucky (O/U)*: Unfavorable performance outcomes that can be explained either by overly optimistic decision making, or by an unlucky turn of events (e.g. a random event that caused demand to be less than predicted).
  - *Pessimistic or Missed opportunity (P/M)*: Unfavorable performance outcomes that can be explained either by overly pessimistic decision making, or by missed opportunities (e.g. an unpredictable or random event led demand to be more than predicted.)
  - *Undetermined Problem*: Unfavorable outcomes that cannot be classified as O/U or P/M, at least not without additional data.

To construct a DDQ tree, we take the set of outcomes and repeatedly split them into finer and finer subsets. At each stage of this process, a performance measure or exogenous variable is chosen, and the outcomes are split based on a comparison between the benchmark and the actual value realized on that measure. If the benchmark for a measure is not binary, but affords a higher level of granularity, then we may have more than two branches emanating from the node. The order in which the measures or exogenous variables are used for splitting depends on three things:

- Which variable provides a more definitive classification of the outcomes at that level? Go from the most discerning to the least discerning variable. This is similar to the idea of entropy reduction used to build decision trees in other contexts [26].
- Are some measures more practically important than others? If so, the more important measures are used before moving to the less important ones.

- The existence of data or cost of obtaining it.

In our case, both Occupancy and ADR are equally important and data is available to permit classification on both measures. However, a split based on Occupancy provides a more definitive classification than one based on ADR, leading to it being used first. We can also use variables other than performance measures for splitting if they increase the diagnostic value of the classification. For example, we used unconstrained demand (an exogenous variable) at the second level of the tree.

The dotted lines on the DDQ tree in Figure 2 show three levels of granularity of analysis. Specifying benchmarks and calculating performance measures lower down in the tree is needed only if that level of granularity is desired (e.g., to more finely diagnose strengths or weaknesses in decision processes). The ability to work at varying degrees of granularity of analysis allows the DDQ analyst to select a level of granularity sensitive to both the costs and benefits of further analysis. The following description proceeds down the tree one level at a time.

*Denials.* As already discussed, the absence of denials (Denials = 0) implies that the outcomes do not reflect the quality of the judgments/decisions made. These outcomes therefore cannot be classified as favorable or unfavorable. If Denials > 0, then the key questions are: Were the denials necessary? Were they made optimally to increase revenue?

*Occupancy.* If the hotel did not fill to capacity, that is, Occupancy < 100%, then the fact that customers were turned away (Denials > 0), clearly indicates that the outcome was suboptimal. A more complete categorization of the outcomes requires further analysis. If Occupancy was 100%, then denials were warranted. However, it is necessary to determine if they were made optimally, that is, were low paying customers turned away so that high paying customers could be accommodated?

*Unconstrained Demand versus Capacity with less than full occupancy.* If Unconstrained Demand > Capacity, the situation is one of excess demand. With excess demand, some denials were clearly necessary, but the lack of full occupancy indicates that denials were excessive. We need further analysis at the next level of the tree to more precisely identify the cause of the unfavorable decision outcome. If Unconstrained Demand < Capacity, then since we have less than full occupancy and a positive number of

denials, the outcome is classified as *Unfavorable – O/U (Optimistic/Unlucky)*. This outcome may have resulted from overly optimistic demand forecasting and/or an overly aggressive pricing strategy given the demand forecast. Alternatively, it may simply be that the random variation in demand led to an unlucky outcome, but the RM's decisions were appropriate for the predictable component of demand. It is not possible to distinguish between these two possibilities without further analysis.

*ADR<sub>actual</sub> versus ADR<sub>market</sub>.* Improving ADR is clearly a desirable outcome. ADR can be improved by turning away low paying customers to accommodate high paying customers, assuming of course that occupancy is not sacrificed in the process. To assess improvements in ADR we compare  $ADR_{actual}$  to  $ADR_{market}$ .  $ADR_{actual}$  is the average daily rate that the hotel actually obtained for the given reservation night. The benchmark  $ADR_{market}$  is our estimate of the ADR that would be obtained for a room night if the sales mix of the hotel matched that of the mix of customers in the market demand curve for that night i.e., the total number of people of different willingnesses-to-pay that seek a room. (The process for estimating the market demand curve was described above.) Where there is full occupancy (the right side of the DDQ tree) and  $ADR_{actual}$  exceeds  $ADR_{market}$  then the outcome is classified as *Favorable* as it meets the twin objectives of 100% occupancy and an improvement in average realized revenue. Otherwise the outcome is classified as *Unfavorable – P/M (Pessimistic/Missed Opportunity)*. Such an outcome may have resulted from overly pessimistic demand forecasting and/or an overly conservative pricing strategy given the demand forecast. Alternatively, it may simply be that random variation in demand led to a lucky outcome of full occupancy. Although the forecasting and pricing in this case may have been appropriate for the predictable component of demand, the RMs failed to capitalize on the unexpected opportunity.

On the left side of the DDQ tree, where there is less than full occupancy, but excess demand, a situation in which  $ADR_{actual} > ADR_{market}$  indicates overly optimistic forecasting, too aggressive pricing or an unlucky random fluctuation in demand resulting in an outcome classification of *Unfavorable-O/U*. Otherwise, the outcome is classified as *Unfavorable-Undetermined problem*.

The construction of the tree explicitly allows for variability in the data and analysis requirements. With the increasingly finer partitions obtained by moving down the tree, fewer of the outcomes fall into the *Undetermined problem* diagnostic class and greater diagnostic power is obtained.

### **The diagnosis phase: Step 1-- Interpreting the results of the DDQ classification**

Analysis of the number of outcome instances falling into each DDQ class provides evidence of the extent to which fluctuations in performance are due to systematic effects (such as decision making biases) rather than random variation. The distribution of outcomes classified as P/M versus O/U should appear random – unless there is a systematic effect (assuming a sufficient number of outcomes in the data set). Thus, the results of the DDQ classification provide information akin to the  $G$  measure of the lens model, in that it facilitates the separation of random from systematic causes of performance fluctuation. Recall however that providing  $G$  is insufficient for learning.  $G$ , per se provides no guidance as to how performance can be improved -- hence the need for further analysis of the DDQ tree classifications in the diagnosis phase.

Figures 3a and 3b show the DDQ trees for “Big City” and “Southern Country” respectively. Immediately we see the value of the DDQ approach. Both trees show a majority of outcomes falling into the *Unfavorable-O/U* category<sup>1</sup>. While some of these outcomes may have arisen due to bad draws from a stochastic market demand, the high number of outcomes in this category clearly suggests a systematic bias in the decision making towards overly optimistic forecasting and/or overly aggressive pricing. But what drives this bias? One obvious possibility is that this is a function of the economic incentives of the task. In the hotel context, an overly optimistic bias towards, for example, aggressive pricing is warranted only if the expected gain in revenue derived from the possible sale to a high rate customer exceeds the revenue forgone from the more certain low rate customer turned away. In Alpha hotels the lowest rate in the pricing scheme averages nearly twice the difference between the lowest rate and the highest rate and,

---

<sup>1</sup> At first glance, the preponderance of unfavorable outcomes may seem implausible, particularly given the economic implications for the hotel. However, recognize that the tag “unfavorable” doesn’t indicate that the decisions involved gross errors – they merely indicate that the outcomes were suboptimal and there is scope for improvement if the appropriate drivers are identified.

consequently, if anything, the economic incentive for the hotel is to be pessimistic in forecasting and conservative in pricing. Nevertheless, the RMs were too optimistic -- a bias we explore further later.

-----  
Insert Figures 3a and 3b Here  
-----

### **The diagnosis phase: Step 2 -- Analysis of outcomes within DDQ classes**

We next attempt to discover the context-specific drivers of the systematic patterns observed in the two DDQ trees for Alpha. We do this by analyzing the decision outcomes within the DDQ classes. While other analytical methods such as regression might also be appropriate for this task, we find it useful<sup>2</sup> to construct control charts in which the decision outcome instances are plotted against a relevant cue variable. It is a long established finding in judgment analysis that human judgment is relatively good at identifying relevant cues, but notoriously bad at weighting cues in a specific judgment [15,24]. The identification of potentially relevant cues is therefore relatively simple when done in consultation with domain decision makers. In any event, the diagnosis phase can be viewed as iterative – the analysis of DDQ charts for some cues sparks investigation of additional cues for analysis, a process that can iterate until relevant task information is discovered.

Control charts are constructed to analyze the data at the leaf or interior nodes and can be done at any level of granularity. For instance, they can be constructed at the macro level of *Favorable*, *Unfavorable* or, within the latter category, further broken down into the *O/U* and *P/M* subcategories. Interpretation of the charts can be aided by plotting mean values and “control limits” based on the standard deviation in the proportions for a given DDQ tree classification across the cue value set.

A DDQ chart of the proportion of *favorable* classifications by cue value can be useful in identifying strengths in decision making processes and thus can aid in determining best practices. A DDQ chart of the proportion of *unfavorable* classifications is diagnostic of systematic deficiencies in the decision processes. Charts at the more micro-level *O/U* and *P/M* classifications provide additional

---

<sup>2</sup> The relative simplicity of their construction and the graphical presentation of information makes control charts much more accessible and appealing to these decision makers.

diagnostic feedback for narrowing in on decision process failures in situations where the macro-level *unfavorable* classification DDQ chart provides insufficient diagnostic information. Since, in Alpha's case, the unfavorable outcomes were almost exclusively falling into the *O/U* class we show plots only of the macro level *Unfavorable* DDQ control charts.

Several cues were obvious candidates over which to plot DDQ charts even before consulting with domain experts: room night, unconstrained demand, and revenue manager experience among them. Room night was clearly the prime candidate as demand for hotel rooms varies from season to season and within days of the week. Therefore the first DDQ control charts we constructed were for day of the week and seasonal fluctuations. Figure 4 shows the DDQ control charts by day of the week for the two Alpha hotels. While the outcome classifications from the DDQ Tree for the two sites looked similar, the DDQ charts suggest that the causes may be quite different.

-----  
Insert Figure 4 Here  
-----

We discussed the DDQ trees and charts with management at Alpha in order to prompt the discovery of task information. The DDQ chart for Big City clearly shows that the RM performance is consistently inferior on Friday and Sunday room nights compared to other week days<sup>3</sup>. Once the situation was stated like this, the RM for Big City almost immediately recognized that Friday and Sunday nights were nights for which they had greater variability in cancellations. Subsequent analysis of cancellation data confirmed the RM's interpretation. The DDQ analysis thus led to the discovery of pertinent task information that difficulties in estimating cancellations were a primary driver of inferior performance. Consequently the Big City RM recognized the need to expend greater effort in estimating cancellations.

The DDQ chart for Southern Country shows that the RM performance at that location was consistently inferior on Thursday and Friday nights<sup>4</sup>. The director of revenue management for the Alpha

---

<sup>3</sup> Statistical tests reveal that the proportion of unfavorable decisions on Fridays & Sundays is significantly higher than those for the rest of the week ( $p = 0.001$ ). However we prefer using the control charts to statistical tests as they are more accessible and intuitive to decision makers.

<sup>4</sup> Difference is statistically significant at the  $p = 0.003$  level.

chain suggested that the RM was using an aggressive pricing strategy throughout the week when it was only appropriate for the first part of the week. He specifically noted that the hotel had a substantial group business in the earlier part of the week, making it easier to fill the hotel with transient business and thus warranting a more aggressive approach. This interpretation is worth noting because it was actually incorrect – analysis of the data indicated that there was no significant difference in the volume of group business between earlier and latter parts of the week. Demand at the highest price point, however, showed greater variability on Thursdays and Fridays as compared with the other days of the week. Consequently we recommended that Southern Country focus its efforts on improving its demand forecasting for the highest rate tier. This experience shows that DDQ trees and charts, while invaluable tools in guiding the discovery process, are by themselves insufficient in identifying systematic problems with the decision process. The same is true of the domain decision makers’ interpretations. To be useful, the tools and interpretations should be combined with some disciplined data analysis to verify the insights developed.

It is important to note that the DDQ control chart provides substantial diagnostic information over and above that which could be obtained by simply plotting outcome variables over relevant cue values and attempting to observe systematic patterns. Since outcome variables are a product of both endogenous and exogenous factors, the noise in a simple plot of outcome variables is less likely to produce interpretable patterns. For example, this is well illustrated in Figure 5 in which we plot RevPAR by day of the week for Southern Country (recall that RevPAR is considered the best single measure of performance in the hotel industry). The RevPAR chart suggests that the performance on Thursdays is the best in the week and that Southern country in fact has problems on weekends. This pattern is in reality attributable to the market demand curve for Southern country, where business is generally slow on the weekends, and therefore not related to any systematic variations in the RM’s decision performance. In contrast, the DDQ chart in Figure 4b shows that decision making at Southern Country was poorer on Thursdays and Fridays when RevPar was at its highest.

-----  
Insert Figure 5 Here  
-----

### 3.3 Discussion of the Results from use of the Methodology

Table 3 summarizes the DDQ analysis process for revenue management at Alpha. Recall that the analysis has two goals. The first goal is to assess the decision performance by separating endogenous drivers (decision processes) of performance from exogenous drivers (environmental stochasticity). The second goal is to provide actionable feedback to the decision maker in terms of functional validity information (systematic biases) and task information that can lead to improved performance through appropriate revision of decision processes.

-----  
Insert Table 3 Here  
-----

The first of these objectives is fulfilled in the evaluation phase of the analysis. Following the lens model structure, we first identify a set of criteria ( $Y_e$  - occupancy and ADR) and specify benchmarks (unconstrained demand,  $ADR_{\text{market}}$ ) on these criteria to measure performance ( $r_a$ ). Unlike the laboratory studies of the lens model, where the actual value of the criterion is controlled by the researcher (and is hence known to him/her), here we only observe outcomes that are a function of both the environment and the decision maker's choices. Therefore care should be exercised in selecting appropriate benchmarks for the performance measures such that endogenous drivers of performance can be separated from exogenous ones. In the revenue management context, this is facilitated by the fact that optimal choices for a given set of exogenous factors can be identified ex-post, and these can then be compared to the RM's decisions to arrive at a measure of endogenous performance.

Phase 2 of the analysis, the diagnosis phase, then sought to understand the drivers of performance. An analysis of the distribution of unfavorable outcomes in step 1 revealed a non-random pattern indicating the presence of an aggressive bias in the revenue management decision making. This meant that the RM's judgments were not accurately calibrated to the environmental processes generating the criterion values (functional validity information). Further analysis using DDQ control charts in step 2 identified patterns in favorable and unfavorable outcomes as they related to the day of the week (a cue).

Some introspection and data analysis then revealed the deeper drivers of these patterns (task information - cancellations and variability in premium demand in the two cases respectively). Thus the analysis identified context-specific drivers of sub-optimal performance which, in turn, suggested the appropriate interventions for improvement (e.g. introduce tools for forecasting cancellations at Big City).

Overall, the analysis demonstrated that there was considerable opportunity for improvement in the revenue management decision making processes of the two hotels we studied – the two hotels that Alpha management deemed the best performing hotels in the chain. It was surprising to both us as researchers, and certainly to Alpha management, to find such compelling evidence of overly optimistic/aggressive decision making in forecasting and pricing. This was particularly surprising given that the systematic bias was counter to the economic incentives for the hotel.

The extent of the bias is likely indicative of a more general behavioral problem. It can be interpreted as a failure to recognize “doing nothing” – inaction – as a decision alternative. Specifically, the RMs often closed price buckets (took action) when they would have been better off leaving all price buckets open (doing nothing). This is consistent with the “inaction effect” found by Zeelenberg and colleagues [31] namely that given a series of negative outcomes, greater regret is experienced from inaction as opposed to action. Since, psychologically, inaction is different from action [20, 22], this may lead to “action for the sake of action” as a means of minimizing potential future regret. Added to this, performance evaluation (by the RM or their supervisor) may fail to recognize that good revenue management decision making is about choosing the right time to act (i.e., close price buckets off), and the right time not to act (i.e., leave price buckets open), thereby strengthening the bias. Prescriptively, this suggests that there may be value in efforts to reframe inaction (leaving price buckets open) as an active choice (e.g., redesigning systems to confront RMs regularly with the choice to open or close price buckets, as opposed to defaulting to the status quo). In terms of the value of our methodology, the preceding analysis demonstrates its ability to identify theoretically interesting biases and systematic deficiencies in real world decision making. As such, we see the DDQ methodology as an academically valuable tool for field decision research, as well as a practical tool for decision improvement.

#### **4. CONCLUSION AND IMPLICATIONS**

In this paper we develop a pragmatic methodology for the diagnosis of decision quality in complex real world settings. The objectives of the methodology are two-fold; first, to identify systematic strengths and weaknesses in the decision making process by separating the exogenous and endogenous drivers of performance; and second, to discover context-specific task information that is likely to result in improved decision performance. The effectiveness of the methodology in meeting these two objectives arises from the insight that while the optimal decisions may be difficult to identify ex-ante, they are much easier to specify ex-post. So, as long as sufficient historical data on outcomes and their associated cues is available, one should be able to use the DDQ methodology in the manner illustrated in this paper. Key strengths of the approach are that it specifically takes account of the costs of data acquisition and analysis and that it can handle multiple conflicting objectives simultaneously.

We demonstrated the usefulness of the methodology in the context of hotel revenue management, where we found systematic biases in judgment that led to substantial revenue losses and identified relevant task information that can potentially lead to improved RM performance. Furthermore, the methodology was able to generate specific recommendations for decision support interventions to improve performance. In this context, it quickly became apparent that while the DDQ trees and charts were invaluable tools in analyzing the problem and guiding the discovery process, active collaboration from domain experts and decision makers was absolutely critical to discover the relevant task information. First, we consulted with them to create a list of cues that potentially have an impact on the outcome. Second, we drew on their expertise to interpret the charts, identify systematic drivers of unfavorable performance, and find likely improvements. Because of the repetitive nature of our decision context, once most of the cues have been identified and drivers of poor performance and likely fixes have been articulated, the decision makers themselves, with the help of an appropriate decision support system, can draw on this “codified” knowledge to discover context-specific task information without further

involvement of the experts.<sup>5</sup> However, for the decision maker to adopt and utilize such a system, the system must be understandable to the decision maker. In this context, the simplicity of the DDQ methodology is a key strength in increasing face validity and encouraging acceptance and involvement.

The reader will note a number of similarities between aspects of our methodology and other available tools. Our partitioning of outcomes into different classes (favorable, unfavorable, etc.) using the DDQ tree is similar to classification tasks in the machine learning literature [26]. However, while the machine learning algorithms use tagged outcomes to automatically learn trees that minimize classification error, we do the opposite – we use domain knowledge of the outcomes and benchmarks to develop the tree, so as to make the classes diagnostic of underlying decision biases. The control charts we plot in the diagnosis phase are similarly inspired by their counterparts in the statistical process control literature. However, instead of using the charts to detect a process out of control, we use them as a visual tool to correlate outcome classes to cues, thereby facilitating the discovery of pertinent task information. In fact, as we suggested earlier, DDQ analysts could use other tools such as regression, discriminant analysis, etc., but we recommend control charts due to their visual appeal and ease of use. The primary contribution of our methodology therefore comes, not from the individual tools per se, but the way in which these tools are combined to perform a systematic post-hoc analysis of decision outcomes.

While we used hotel revenue management to provide a proof of concept, the methodology is equally useful in other repeated decision making contexts that involve human judgment. The only requirements are the availability of relevant cue and outcome data ex-post, and the ability to identify the optimal decisions in hindsight based on this data. Take for instance, the management of product inventory in a retail store. A potentially useful outcome variable in this case could be the daily closing inventory. The inventory manager is faced with two conflicting objectives – first, to avoid stockouts, and second, to minimize inventory costs. In arriving at the optimal inventory policy, the manager has to make judgments about demand and order lead times. Since perfect information on both demand and ordering is available ex-post, we can draw a DDQ tree to classify outcomes as favorable and unfavorable at different

---

<sup>5</sup> We thank a reviewer for pointing this out.

levels of granularity (Figure 6 provides a self-explanatory DDQ classification tree for this application). Further we can construct control charts of these classes against cues such as product class, time of the week/month, seasonality, promotions for the product and competing products, etc., to discover relevant task information. Other managerial decision contexts where our methodology can be readily applied include production scheduling, trading (commodities, stocks, currencies or other financial instruments), and repeated pricing and bidding for projects. DDQ analysis can also be used to improve the performance of decision support systems by creating a dynamic learning environment in which DSS users continually refine their models based on task information discovered from an analysis of the data generated in the process of using the DSS.

Although it is not the focus of our research, the DDQ methodology can also be used to understand real world human decision making from a theoretical perspective. Psychologically grounded *field studies* of decision making, are substantially underrepresented in the literature, especially when compared to the plethora of laboratory studies (there are a few notable exceptions, e.g., [25]). Albeit in a small way, our work helps to redress this imbalance. This paper also addresses another imbalance in the literature – namely the under-representation of prescriptive decision making research, relative to normative or descriptive research [3]. The decision making literature abounds with descriptive studies of how people make decisions, but research exploring interventions to improve decision making, informed by those descriptive theories, are not common (again there are a few notable exceptions, e.g., [5]).

## REFERENCES

- [1] W.K. Balzer, M.E. Doherty and R. O'Connor, Effects of cognitive feedback on performance, *Psychological Bulletin* 106(3) (1989), 410-433.
- [2] W.K. Balzer, L.M. Sulsky, L.B. Hammer, and K.E. Sumner, Task information, cognitive information, or functional validity information: Which components of cognitive feedback affect performance? *Organizational Behavior and Human Decision Processes* 53(1) (1992), 35-54.
- [3] D.E. Bell, H. Raiffa and A. Tversky, Descriptive, normative, and prescriptive interactions in decision making, in: D.E. Bell, H. Raiffa, and A. Tversky, Eds., *Decision making: Descriptive, normative, and prescriptive interactions*, Ch. 1 (Cambridge University Press, Cambridge, 1988), 9-32.

- [4] B. Brehmer, In one word: Not from experience, *Acta-Psychologica* 45(1) (1980), 223-241.
- [5] G.J. Browne, S.P. Curley and P.G. Benson, Evoking information in probability assessment: Knowledge maps and reasoning-based directed questions, *Management Science* 43(1) (1997), 1-14.
- [6] E. Brunswik, *Perception and the Representative Design of Experiments*, (University of California Press, Berkeley, CA, 1956).
- [7] C.F. Camerer, and E.J. Johnson, The process-performance paradox in expert judgment. How can experts know so much and predict so badly? In: K. Ericsson and J. Smith, Eds., *Toward a general theory of expertise - prospects and limits*, Ch. 8, (Cambridge University Press, New York, 1991).
- [8] N.J. Castellan, Relations between linear models: Implications for the lens model, *Organizational Behavior and Human Decision Processes*, 51(3) (1992), 364-381.
- [9] R.W. Cooksey, *Judgment analysis: Theory, methods & applications*, (Acad. Press, San Diego, 1996).
- [10] R.G. Cross, *Revenue management: Hard-core tactics for market domination*, (Broadway Books, New York, 1997).
- [11] M.J. Davern, *Performance with Information Technology: Individual fit and reliance on Decision Support Systems*, Unpublished Ph.D. Dissertation, Information and Decision Sciences, Carlson School of Management, University of Minnesota, Minneapolis, MN, (1998), 142.
- [12] M.J. Davern and R. J. Kauffman, *On the Limits to Value of Decision Technologies*, Presentation made at the Tenth Workshop on Information Systems Economics, New York, NY (1998).
- [13] F.D. Davis and J.E. Kottemann, User perceptions of decision support effectiveness: Two production planning experiments, *Decision Sciences* 25(1) (1994), 57-78.
- [14] F.D. Davis and J.E. Kottemann, Determinants of decision rule use in a production planning task, *Organizational Behavior and Human Decision Processes*, 63(2) (1995), 145-157.
- [15] R. Dawes, The robust beauty of improper linear models in decision making, *American-Psychologist*, 34(7) (1979), 571-582.
- [16] R. Fildes, P. Goodwin, and M. Lawrence, The design features of forecasting support systems and their effectiveness, *Decision Support Systems* 42(1) (2006), 351-361.
- [17] K.R. Hammond, G.H. McClelland and J. Mumpower, *Human judgment and decision making*, (Praeger, New York, 1980)
- [18] K.R. Hammond, T.R. Stewart, B. Brehmer and D.O. Steinman, Social judgment theory, in: M. F. Kaplan and S. Schwartz, Eds., *Human judgment and decision processes* (Academic Press, New York, 1975), 271-312.
- [19] A.R. Hevner, S.T. March, J. Park, and S. Ram, Design Science in Information Systems Research, *Management Information Systems Quarterly* 28(1) (2004), 75-105.
- [20] D. Kahneman and A. Tversky, The Psychology of preferences. *Scientific American* 246(1) (1982), 160-173.

- [21] G.W. Keren and B. deBruin, On the assessment of decision quality: Considerations regarding utility, conflict and accountability, in D. Hardman and L. Macchi, Eds., *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making*. (John Wiley, Chichester, 2004).
- [22] J. Landman, Regret and elation following action and inaction: Affective responses to positive versus negative outcomes, *Personality and Social Psychology Bulletin* 13(4) (1987), 524-536.
- [23] S.T March and G.F. Smith, Design and natural science research on information technology, *Decision Support Systems* 15(4) (1995), 251-266.
- [24] P.E. Meehl, *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*, (University of Minnesota Press, Minneapolis, 1954).
- [25] N.P. Melone, Reasoning in the executive suite: The influence of role/experience-based expertise on decision processes of corporate executives, *Organization Science* 5(3) (1994), 438-455.
- [26] T.M. Mitchell, *Machine Learning* (McGraw-Hill, New York, 1997).
- [27] P. Slovic and S. Lichtenstein, Comparison of bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance* 6(6) (1971), 649-744.
- [28] F.J. Todd and K.R. Hammond, Differential feedback in two multiple-cue probability learning tasks, *Behavioral Science* 10(4) (1965), 429-435.
- [29] L.R. Tucker, A suggested alternative formulation in the development of Hursch, Hammond and Hursch and by Hammond, Hursch and Todd, *Psychological Review* 71(6) (1964), 528-530.
- [30] J.F. Yates, *Judgment and decision making* (Prentice Hall, Englewood Cliffs, NJ, 1990).
- [31] M. Zeelenberg, K van den Bos, E. van Dijk, and R. Pieters, The inaction effect in the psychology of regret, *Journal of Personality and Social Psychology* 82(3) (2002), 314-327.

**Table 1: Types of cognitive feedback**

<b>TYPE OF FEEDBACK</b>	<b>DEFINITION AND MEASURES</b>
Task Information (TI)	Information about the task environment (e.g., $r_{ei}$ , $r_{ij}$ , $R_e$ )
Cognitive Information (CI)	Information about the subject's own cognitive system (e.g., $r_{si}$ , $R_s$ )
Functional Validity Information (FVI)	Information about the relationship between the subject's cognitive system and the task environment (e.g., $r_a$ , $G$ ).

**Table 2: The generic Diagnosing Decision Quality methodology**

<b>PHASE</b>	<b>DIAGNOSTIC STEP</b>	<b>LENS MODEL "EQUIVALENT"</b>
<b>Evaluation</b>	1. Identify a portfolio of performance measures, corresponding benchmarks and exogenous variables.	$Y_e$ , $r_a$
	2. Construct a DDQ classification tree to facilitate the separation of systematic decision effects from exogenous/random outcome fluctuations. Classify decision outcomes.	$R_e$ versus $G$
<b>Diagnosis</b>	1. Interpret the results of the DDQ classification process by analyzing the distribution of decision outcomes across diagnostic classes in the DDQ tree.	Functional Validity Information (FVI): $G$
	2. Analyze outcomes within the diagnostic classes by identifying potentially relevant cue variables and constructing control charts to prompt discovery of causal explanations of systematic patterns.	Task Information (TI): $X_i$ , $r_{ei}$

**Table 3: Summary of DDQ analysis for hotel revenue management**

<b>PHASE</b>	<b>DIAGNOSTIC STEP</b>	<b>HOTEL REVENUE MANAGEMENT CONTEXT</b>
<b>Evaluation</b>	1. Identify a portfolio of performance measures, corresponding benchmarks and exogenous variables.	$Y_e$ - Measures: Occupancy, ADR $r_a$ - Benchmarks: Full Occupancy, ADR <sub>market</sub> Exogenous variables: Unconstrained demand
	2. Construct a DDQ classification tree to facilitate the separation of systematic decision effects from exogenous/random outcome fluctuations. Classify decision outcomes.	Classes: <ul style="list-style-type: none"> <li>- Favorable</li> <li>- Unfavorable – Optimistic/Unlucky</li> <li>- Unfavorable – Pessimistic/Missed Opportunity</li> <li>- Unfavorable – Undetermined Problem</li> </ul>
<b>Diagnosis</b>	1. Interpret the results of the DDQ classification process by analyzing the distribution of decision outcomes across diagnostic classes in the DDQ tree.	Functional Validity Information (FVI): $G$  -- Overly optimistic demand forecasting/ aggressive pricing by RMs
	2. Analyze outcomes within the diagnostic classes by identifying potentially relevant cue variables and constructing control charts to prompt discovery of causal explanations of systematic patterns.	Task Information (TI): $X_i$ - Day of week, season (cue variable) -- Poor understanding of cancellations in the Big City hotel. -- Poor demand forecast for the upper price segment in Southern Country hotel.

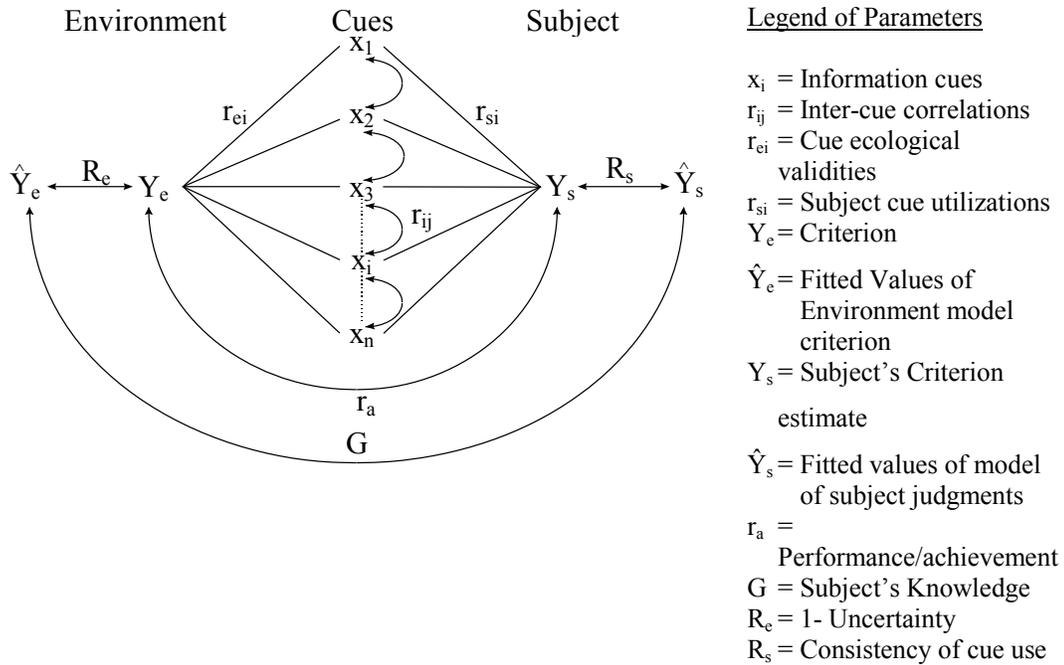


Figure 1 - The Lens Model

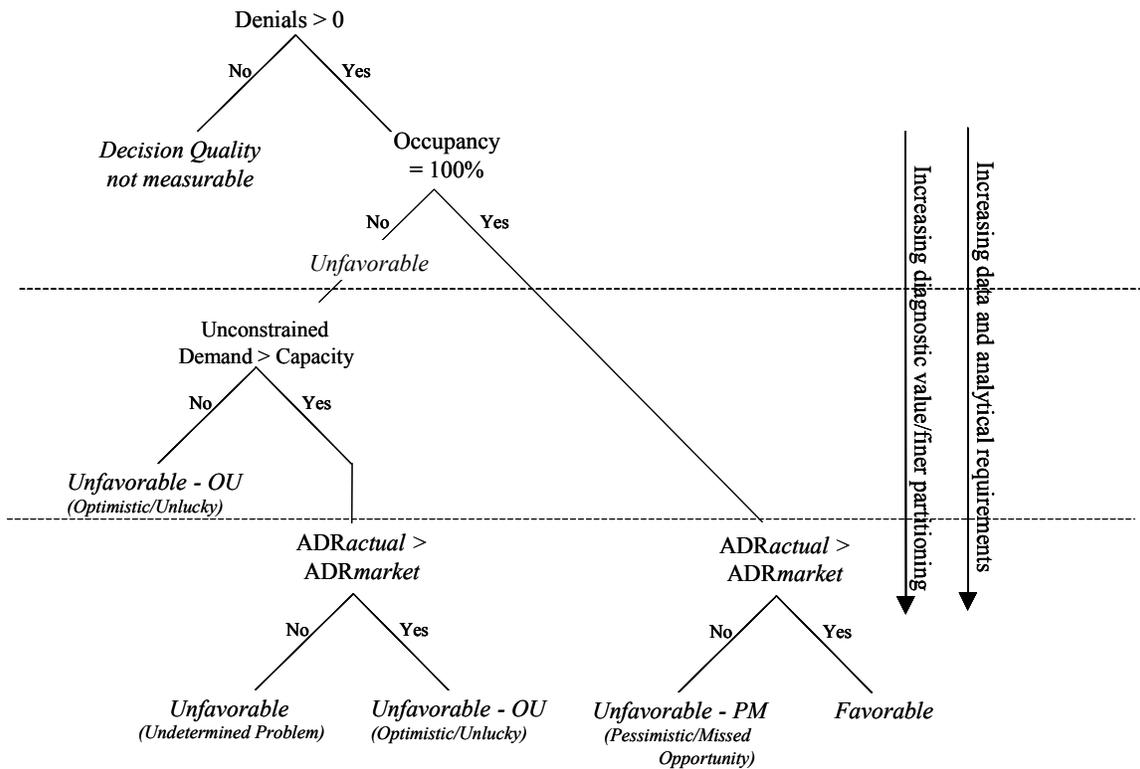
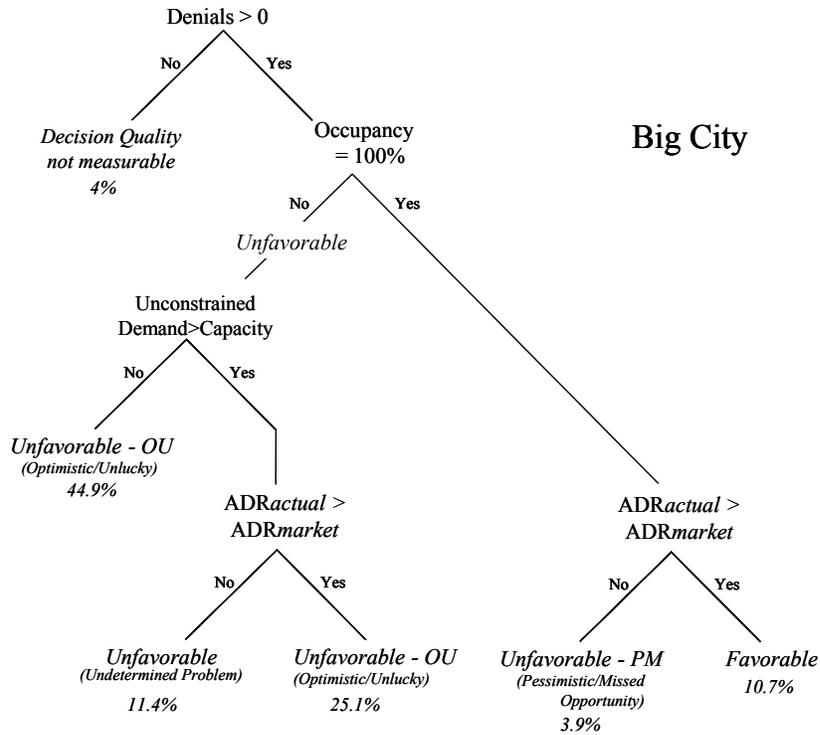
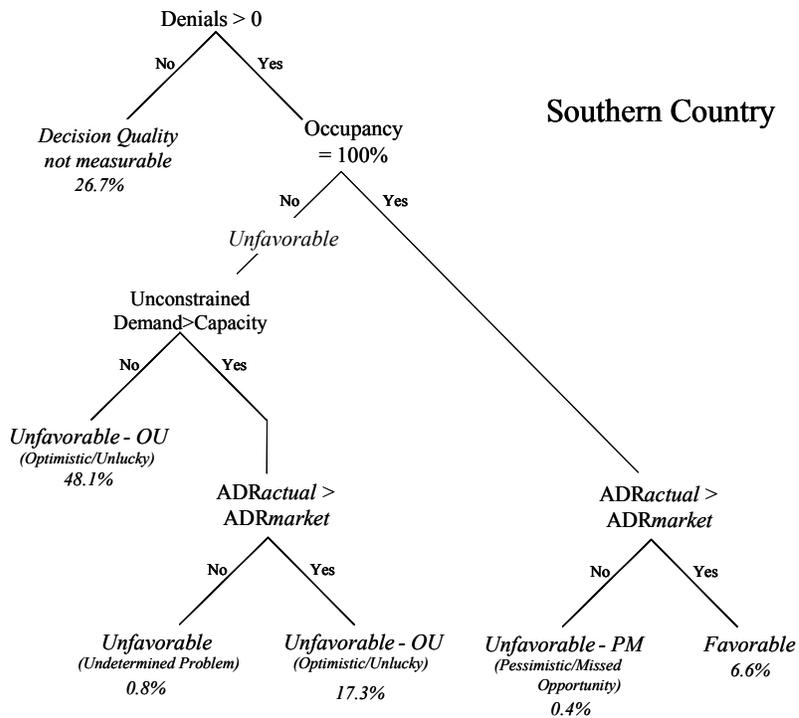


Figure 2: DDQ Tree for hotel revenue management



**Figure 3a: DDQ Tree for Alpha Hotels – Big City**



**Figure 3b: DDQ Tree for Alpha Hotels – Southern Country**

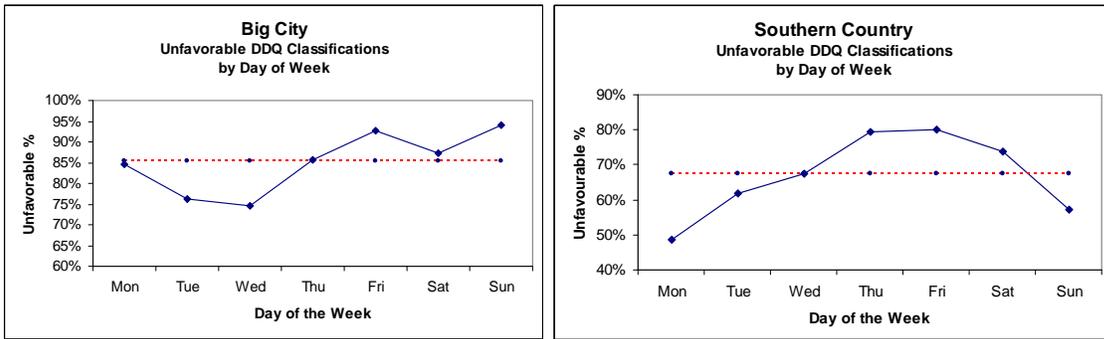


Figure 4: DDQ Control charts for Alpha Hotels

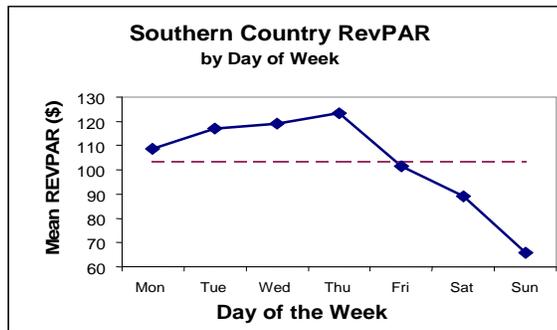


Figure 5: RevPAR by day-of-week for Southern Country

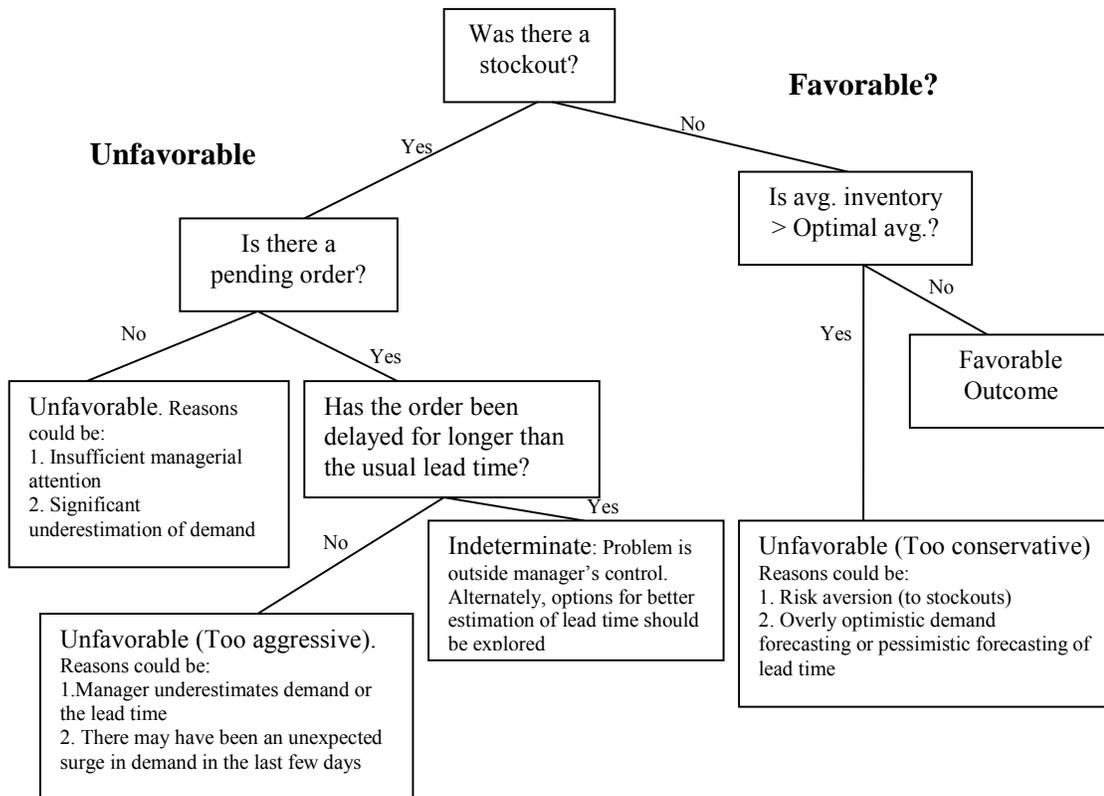


Figure 6: DDQ Tree for retail inventory application