Human decision-making has often being criticized as boundedly rational, and subject to biases (e.g., Tversky and Kahneman 1974). Important information cues are often misweighted (Dawes 1979, Camerer & Johnson 1991) and feedback delays hamper learning (Sterman 1989). 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 (Goodwin and Fildes 1999, Davern and Kauffman 1998, Davern 1998, Davis and Kotteman 1995, 1994). Experimentally information such as cue-criterion relationships has been shown to be of incremental value (Balzer et al. 1992, Balzer et al. 1989). In the field such information is only readily available in trivial settings. Generalizing from lab based approaches, we present a new field methodology for the “Diagnosis of Decision Quality” (DDQ) that aids decision makers in the discovery of such diagnostic information. We illustrate the application of our approach in the context of hotel revenue yield management using daily data obtained from an international upper-upscale hotel chain. We demonstrate the effectiveness of our methodology by identifying context specific systematic failures in decision making processes in a manner that facilitates the generation of adaptive changes (e.g., the introduction of specific decision aids) in the process to improve subsequent performance. More generally, we provide evidence of a decision making bias consistent with a failure to recognize “doing nothing” as a valid course of action. Significantly, this bias is consistent with the rhetoric of revenue management but is counter to the economic incentives of the revenue manager and hotel chain.

© Copyright 2002, Davern, Mantena & Stohr. Comments welcome, please do not quote, cite or redistribute without permission.
1. Introduction

The substantial literature in judgment analysis (Cooksey 1996) has ably demonstrated that improving human decision making processes is a non-trivial exercise. Humans, while adaptive, also misinterpret feedback and inappropriately weight relevant information cues (Dawes 1979, Camerer & Johnson 1991). 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 (Goodwin and Fildes 1999, Davern and Kauffman 1998, Davern 1998, Davis and Kotteman 1995, 1994). What decision makers require is diagnostic information and feedback – information that can assist them directly in revising their decision processes rather than simply reporting the outcomes. In laboratory settings, the provision of task information, such as details of cue-criterion relationships, has been shown to be key to improving performance in judgment tasks (Balzer et al 1992). In rich and real organizational settings however, such information is obviously difficult to obtain for all but trivial tasks (indeed it is the lack of information about cue-criterion relationships that makes many real world tasks so difficult.)

In this paper we present a new methodology for the “Diagnosis of Decision Quality” (DDQ) in the field, with a particular focus on the context of hotel revenue management decision making. We base this method on an analysis of the generic diagnostic characteristics of the well known lens model (Brunswik 1956) that is the basis of decades of laboratory-based judgment analysis studies (Cooksey 1996, Balzer et al., 1989). We show how these diagnostic characteristics can be mimicked in the field setting using our methodology. We illustrate the application of our approach in the context of hotel revenue yield management using daily data obtained from an international upper-upscale hotel chain.

The structure of the paper is as follows. First, we provide an analysis of the defining characteristics of laboratory-based judgment analysis methods as a guide for the development of new judgment analysis approaches (both field-based, such as ours, or lab-based.) Second we present our methodology and provide a proof of concept (March & Smith 1995) in its application to hotel revenue yield management. We also demonstrate the practical value of our methodology by identifying context-specific systematic failures in decision making processes in a manner that facilitates the generation of adaptive changes (e.g., decision support interventions) in the process to improve subsequent performance. Finally, we provide some reflections on the use of our methodology and the field evidence it provides of a psychologically interesting bias in decision
making that is counter to the real world economic incentives inherent in the revenue management task.

2. Theoretical Background and Development:

2.1 Feedback and 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 (1990, 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 (Slovic & Litchenstein 1971). The lens model, which employs regression, has been used extensively in laboratory studies of human judgment for some forty years, originating with the work of Brunswik (1956). The fundamental philosophy of lens model studies of individual judgment is that to fully understand 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 and hence we build on this tradition and frame our theoretical development in terms of the lens model. Figure 1 shows a common formulation of the lens model.

Under the lens model, the environment is modeled as a linear regression of a set of information cues (xi), which may be correlated (ri) and that are predictive (rei) of some criterion variable (Ye) (e.g., temperature, a cue, might be predictive of the criterion variable future sales of ice cream). The correlation (rei) between a cue (xi) and the criterion (Ye) is referred to as the ecological or predictive validity of the cue. Because the environment is stochastic, this linear model of the environment is not perfectly predictive. The extent of uncertainty or randomness in the environment is indicated by the correlation (Rε) between the actual outcome (Ye) and the predicted outcome (Ye) (Rε=1 implies a certain and perfectly linear environment). Of course, in an experimental setting, the environment model is known and controlled by the experimenter.
The decision maker or experimental subject responds to multiple sets of cues with varying values and, in each case, makes a prediction ($Y_s$) of the criterion. Using regression a linear model for each subject is constructed describing the weight ($r_{si}$), on average, applied to each cue by the subject in predicting the criterion. The weight ($r_{si}$) applied to a cue 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 (Hammond 1972; Hammond et al. 1975). $G$ is essentially a measure of how well a subject’s cue utilizations correspond to the predictive validity of the cues – how appropriate the weights are that the subject gives to the cues. Judgment performance is given by the lens model equation (Tucker 1964) a function of knowledge ($G$), a subject’s consistency in applying his knowledge ($R_s$) and environmental uncertainty (1-$R_e$):

$$r_a = GR_s R_e \text{ (assuming uncorrelated residuals)}$$

---

1 Uncorrelated residuals require that the environment be linear. While this constraint is potentially unrealistic outside the lab, it is “designed in” by experimenters in the construction of the experimental task.
Types of Feedback. Simple outcome feedback (e.g. in lens model terms $Y_e$ or $r_e$) while the most readily available feedback information, both in the lab and in the field, is also the least valuable. Indeed, in a review of several studies Brehmer (1981) presents evidence that the outcome feedback can actually be harmful to learning. Outcome feedback is non-diagnostic, telling the decision maker only the quality of the results and nothing about how to adapt decision processes to improve performance. More formally this is evident in the lens model equation. Given unfavorable performance outcomes (low $r_e$), it is impossible to distinguish the cause of the failure among the competing alternatives; weak knowledge ($G$), inconsistency in the application of the knowledge (1-$R_e$), or the unpredictability of the environment (1-$R_e$).

An alternative to simple outcome feedback is Cognitive Feedback. Balzer et al. (1989) 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).

Table 1 – Types of Cognitive Feedback

<table>
<thead>
<tr>
<th>Type of Feedback</th>
<th>Definition and measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task Information (TI)</td>
<td>Information about the task environment (e.g., $r_{ei}$'s, $r_{fj}$, $R_e$)</td>
</tr>
<tr>
<td>Cognitive Information (CI)</td>
<td>Information about the subject’s own cognitive system (e.g., $r_{si}$, $R_s$)</td>
</tr>
<tr>
<td>Functional Validity Information (FVI)</td>
<td>Information about the relationship between the subject’s cognitive system and the task environment (e.g., $r_a$, $G$).</td>
</tr>
</tbody>
</table>

In a comprehensive review of studies employing the lens model, Balzer et al (1989) conclude that Task Information appears to be the most useful. In a later study, Balzer et al (1992) empirically compared five different feedback conditions (TI only, CI only, TI+CI, TI+CI+FVI, and no feedback). Task information was again found to lead to significant improvements in performance, cognitive information alone never outperformed the no feedback, and the addition of CI and FVI on top of task information did not lead to any additional performance gains.

---

2 Outcome feedback has been found to assist learning only in the simple situations with two or fewer cues linearly related to a criterion (Cooksey 1996).
Experimentally, Task Information is the most useful aspect of cognitive feedback for learning. It provides direct guidance in adapting existing decision processes. In practice however, TI is difficult to obtain. Indeed if the cue-criterion and inter-cue relationships are well known then the judgment task becomes quite trivial and can be readily automated. In the field the cue sets are large, the intercorrelations often complex, and cue-criterion relations may be non-linear. Difficult tasks, those that are not readily automated, those that require the exercise of human judgment, are difficult largely because task information is not easy or economical to obtain.

What real world decision makers require is diagnostic feedback that helps them discover relevant task information about 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. 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 methodology described below addresses some of these concerns.

2.2. From the Lab to the Field: Characteristics of a Methodology for Diagnosing Decision Quality

Theoretical Characteristics. There are several key characteristics of the lens model that underlie its usefulness in analyzing judgments and building a field-based analysis methodology for diagnosing decision quality. Specifically:

1. A criterion against which performance may be assessed ($Y_e$)
2. Separation of random variation in performance from systematic bias (i.e., poor $G$)
3. Identification of relevant cues in the environment ($X_i$)
4. 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 – $r_{si}$).
Items 1 and 2 are essential for the evaluation of judgment performance from an outcome point of view. Items 3 and 4 are essential for an analysis of the knowledge and learning of the decision maker – as they provide a model of the environment for comparison with the decision maker’s behavior. In terms of the diagnosis of decision quality, items 1 and 2 permit the classification of judgments as “good” or “bad”, as distinct from good and bad outcomes. Items 3 and 4 are diagnostic about what was “good” or “bad” about the judgments that have been made.

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 bias and 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 (e.g., inappropriate r_i’s) or misvalued and are the primary cause of unfavorable performance evaluations.

Pragmatic Constraints. In addition to the inherent difficulties of achieving each of these steps in the field (e.g., simply obtaining an ideal performance benchmark can be non-trivial) several other pragmatic constraints can hinder the development of an effective field methodology for the diagnosis of decision quality. Specifically:

1. Multiple objectives imply the need for performance measures for each objective. Any methodology must recognize the multiple competing objectives that are typical of real world decision situations and consequently a decision makers’ performance needs to be evaluated in terms of multiple measures.

2. Data availability and/or acquisition costs. A good methodology should allow for varying depths of analysis, conformant with the availability of data and the cost of acquisition of the data for the diagnosis of decision quality.

3. 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.

---

3 Stenson (1974) describes an approach for estimating lens model parameters such as R_c, R_s, and G, without explicit information about cues and cue-criterion relationships. However, the method requires measuring judgment and criterion values at two points in time, under essentially identical conditions. Such a constraint is hard to realize in the field, both because field settings can be complex and dynamic, and because obtaining identical conditions mandates that no decision feedback be provided between the two measurement points. Consequently, as Cooksey (1996) notes, Stenson’s approach has rarely been utilized.

4 In real world settings, the input cues for some judgments are themselves the products of judgment, which itself may be subject to misjudgment.
2.3 An Overview of the Diagnosing Decision Quality Methodology.

The development of our DDQ methodology follows the four-step pathway outlined above, while remaining cognizant of the pragmatic constraints. Generically it can be described in two phases of two steps each (see Table 2). In what follows we provide a brief overview of the principles of the methodology and then in Section 3 provide a more detailed description of the approach and illustrate its application in the hotel revenue management context.

Table 2 – The Generic Diagnosing Decision Quality Methodology

<table>
<thead>
<tr>
<th>Phase</th>
<th>Diagnostic Step</th>
<th>Lens Model “Equivalent”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation</td>
<td>1. Identify a portfolio of performance measures and corresponding criterion or benchmark.</td>
<td>$Y_{e}, r_a$</td>
</tr>
<tr>
<td></td>
<td>2. Construct a DDQ classification tree to facilitate the separation of random outcome fluctuations from systematic effects.</td>
<td>$G$</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>3. Identify potentially relevant cue variables that may lead to the explanation of the systematic effects observed in the DDQ tree</td>
<td>$X_i$</td>
</tr>
<tr>
<td></td>
<td>4. Construct control charts of DDQ tree classifications by cue values to prompt discovery of causal explanations of the systematic patterns in the outcome classifications from the DDQ tree.</td>
<td>$r_{ei}, r_{si}$</td>
</tr>
</tbody>
</table>

**The Evaluation Phase.** The purpose of the evaluation phase is to separate out fluctuations in performance due to random effects from those due to systematic biases in the judgment and decision making processes employed. In essence, we seek a field equivalent to the $G$ measure of the lens model lab studies.

The first step in evaluation is the establishment of performance measures and benchmarks. Since in the field performance is multi-attribute (due to multiple objectives), a portfolio of measures is likely. Good integrative measures reflecting the attainment of multiple objectives, are typically rare. For each of these performance attributes, a benchmark or criterion needs to be defined. Importantly at this step in the methodology, it is not essential that all the measures and benchmarks be readily calculable or even estimable from available data. Indeed, as shown below, this is a critical strength of our approach – measures and benchmarks are determined only when diagnostically necessary.
The second step in the evaluation phase is to develop a minimalist classification tree (the DDQ Tree) that takes as input the portfolio of the performance variables and corresponding criterion values for each decision outcome and, at the simplest level, classifies them as favorable or unfavorable. The result is a set of decision outcome classifications that facilitate the separation of performance fluctuations due to random variation from those due to systematic effects (strengths or weaknesses in underlying decision making processes, or systematic variations in task difficulty). Systematic effects can be deduced from the distribution patterns of the classifications. 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 some 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.

**The Diagnosis Phase.** The diagnosis phase provides a more elaborate analysis of the systematic patterns in the DDQ classifications, with a view to discovering the systematic causes for any observed patterns. To this end we plot DDQ classifications over decision relevant cue variables to construct DDQ Control Charts.

There are two steps in diagnosis - the identification of relevant cues and DDQ control chart analysis. The involvement of the domain decision makers is critical here, not simply in helping identify relevant cues but also in interpreting the charts to discover systematic causes. Thus, the cues chosen need not in themselves be direct causes of the systematic patterns in DDQ classifications, they merely need to be related to the cause sufficiently 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 leads the decision maker to discover the cues for which the cue utilizations ($r_{c_i}$) deviate most significantly from the predictive validities ($r_{ei}$) (i.e., the cue is misweighted). In practice some cues themselves require a priori judgment, in which case, the discovery process iterates, first by identifying the cue that is poorly judged and then by initiating a process to discover systematic biases in the judgment of the cue value itself.
3. The DDQ Methodology Applied: A Proof of Concept

3.1 The Research Context: Revenue Management at Alpha Hotels

Revenue management (RM) provides an excellent context for illustrating the application of our DDQ methodology, particular as applied in the hotel industry. RM is a recurring task, of substantial economic significance, and while computer-based revenue management systems are becoming increasingly prevalent the end decision still involves a large amount of human judgment (Davern 1998, Inge 1999). The basic tenets of revenue management are market segmentation with multiple price points, and the use of price to handle fluctuations in demand when constrained by a fixed short-term capacity (Cross 1997, Weatherford & Bodily 1992). It is typically applied in situations of high fixed costs, low variable costs and where the opportunity for revenue is perishable. Examples include airlines, hotels, car rental, advertising, and even discount stores like Wal-mart (where the perishable opportunity may not be simply the product (e.g., a seasonal or fashion item), but the shelf space a product for sale occupies). In essence, revenue management entails effective price discrimination to shift consumer surplus back to the supplier. Thus, in times of oversupply, a firm practicing revenue management will increase its low price segment sales. Conversely, in times of shortage, a firm may be better off turning away low price customers in the expectation that a sufficient number of high price segment customers will be forthcoming before the product perishes and the revenue opportunity is lost. Thus revenue management involves two critical judgments, a demand forecast and a price/product availability optimization given that demand forecast. Even its most complex application – the airline industry – these judgments, while supported by computer tools actively involve human judgment (Robertson 1997).

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

Following revenue management practices, hotels offer rooms at a number of price points or buckets, for example (in descending rate), premium, corporate, contract, 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 reservations tend to capacity, the RM may close the cheaper buckets in an effort to filter out only the highest paying customers for the remaining rooms. Essentially the RM seeks to fill the hotel, with the highest paying customers possible, while also trying to ensure that no rooms remain unsold (since all of a room's revenue potential on a given night expires once the night has passed). In theory this entails effective demand forecasting and appropriate optimization of pricing given a forecast. In practice, the demand forecast is not always made explicitly, making it impossible to assess separately the quality of forecast judgments and pricing decisions.

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 will 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., several weeks – reflecting the seasonal nature of demand). 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 "SuperSaver" bucket earlier in the reservation cycle for a given room night than a more conservative RM. Furthermore, hotels, like airlines, regularly overbook in anticipation of cancellations and “no-shows”, which sometimes results in displaced customers being “walked” to alternative accommodations at substantial expense. Senior hotel management is generally involved in setting broad price structures and attitudes to overbooking and “walks”. At the other end of the spectrum, booking agents seek to maximize sales revenue given the open buckets determined by the RMs within the prevailing price structure.
Hotels routinely track at least three different performance measures: Occupancy, Average Daily Rate (ADR), Revenue per Available Room (RevPAR):

- **Occupancy**: Occupancy, often expressed as a percentage of available rooms that are used on a given room night, is a traditional measure of hotel performance. Improvements in occupancy are indicative of good revenue management decision making to the extent they were not due to increases in market demand or obtained by unnecessarily lowering rates and unduly sacrificing revenues.

- **Average Daily Rate (ADR)**: ADR, calculated as rooms revenue divided by rooms sold, is another stalwart of hotel performance measurement. For ADR to be a useful measure of the quality of pricing decisions it must not have been inflated by sacrificing occupancy or by changes in market conditions.

- **Revenue Per Available Room (RevPAR)**: RevPAR, defined as Occupancy x Average Daily Rate, seeks to combine the strengths of ADR and occupancy and to cancel out some, but not all, of the weaknesses. While RevPAR produces a more comprehensive performance measure it cannot be used in comparing performance in different geographical markets. In the hotel industry RevPAR is widely considered the critical performance measure. Most hotel chains also subscribe and provide data to organizations such as Smith Travel Research that provides contributing hotels with information about the “market” RevPAR for a hotel’s identified set of competitors (although an individual competitor hotels RevPAR is not made available).

The above measures of RM performance are primarily evaluative rather than diagnostic. It is difficult at best to connect performance outcomes to decision processes as opposed to random market fluctuations. RevPAR, the holy grail of hotel performance, only provides a market normed measure when used in conjunction with the widely used Smith Travel Research STAR reports, and then only on a monthly rather room night basis – a level of granularity that makes it practically impossible to causally relate performance outcomes to underlying decision processes.

**The Research Site and Data.** We illustrate the application of our approach in the context of hotel revenue yield management using daily level data obtained from an international upper-upscale hotel chain, we will refer to as Alpha Hotels. Organizationally, Alpha is heavily decentralized with a mix of franchised, managed and chain-owned facilities. Each hotel typically comprises several hundred guest rooms, and substantial conference and dining facilities. Individual hotel managers have substantial autonomy, and consequently, a key problem faced by the chain head office was convincing the hotels of the need to spend substantial funds on revenue management decision support (other major players in the hotel industry have been investing heavily in revenue management decision support tools in recent years (Inge 1999)).
Alpha made available to us daily data for all the hotels in the chain for all the standard performance metrics: occupancy by rate class, average daily rate, average length of stay, denials (discussed further below), cancellations, no-shows and walks. Data on individual reservations and timing of reservations in the booking cycle were not available, however channel data was provided (e.g., occupancy by source of reservation: hotel booking, central reservation service, or travel agent.) Data was provided in the form of Excel spreadsheets submitted to the chain head office by the individual hotels. Unfortunately, not all hotels in the chain provided consistent or complete information. Our analysis focuses on the two hotels that the chain identified as their leading performers and for which the data was most complete: “Big City” located in a major US metropolitan area and “Southern Country” located in a southern regional center.

3.2 Applying the methodology

The Evaluation Phase: Step 1 – Performance Benchmarks

The first step of the evaluation phase of our methodology (see Table 2) involves establishing performance benchmarks. While the three hotel performance measures described above are pertinent outcome measures, establishing benchmarks or criterion values requires a more formal description of ideal revenue management decision making. Assuming the RM has perfect information about demand, the hotel revenue management task can be loosely defined as⁵:

Objective: Maximize Revenue
Constraints: Hotel Capacity, the Market Demand Curve
Action: Pricing Strategy selection: specifically opening and closing price buckets

Decision Rules:
IF Unconstrained Demand ≤ Capacity THEN do nothing
IF Unconstrained Demand > Capacity THEN close out price buckets where price is below P*

WHERE
P* is defined such that Demand at P* = Capacity.
Unconstrained Demand is the number of rooms that could be sold if the hotel had infinite capacity.

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 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

⁵ A number of precise mathematical models of revenue management have been presented in the literature, especially with respect to the airline revenue management problem. (Smith, Leimkuhler, & Darrow 1992).
structure is not continuous – the hotel may not have a price bucket with the exact price $P^*$.
Indeed, hotels typically work with a fixed number of price buckets with prices that are fixed in
the short term (they may vary by day of the week, but be consistent from week to week within a
season). In part, this stems from the practical difficulties of continuous and dynamic pricing and,
in part, from very real concerns about consumer acceptance of such price volatility. The revenue
management task is further complicated by the possibility of multiple night stays -- lower rate
customers may be accepted on an excess demand night because of the revenue they provide on
subsequent or preceding nights for which demand is less than capacity.

A key benchmark in revenue management is the Market Demand Curve. While the entire
Market Demand Curve is difficult to estimate, even ex post, it is possible to estimate
unconstrained demand, 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 because of lack of availability of a room at a price bucket they are
willing to accept. Hotel chains routinely, as part of their booking processes, attempt to track
“denials”, although true denials (i.e., customers turned away who actually would have stayed) can
only be estimated rather than exactly measured. Denials represent customers with a lower
willingness to pay who are turned away in expectation of higher paying customers arising later in
the booking cycle. Denials, average daily rate and occupancy data, are adequate for commencing
the second step of the evaluation phase – the construction of the DDQ tree.

**The Evaluation Phase: Step 2 – The DDQ Tree**

There exists a well-established literature describing guidelines for the construction of
classification trees (e.g., in machine learning, one approach involves splitting trees on variables
so as to minimize entropy)\(^7\). In this paper we focus on somewhat different issues, specific to the

\(^6\) More specifically, these are rate-based denials as distinct from room-type denials, which occur when
customers are turned away because they could not obtain the desired physical room type, irrespective of
price. Room-type denials are a product of the physical hotel characteristics, and are not the result of
decisions regarding the opening and closing of price buckets. Contrary to popular consumer belief, price
buckets and physical room types do not map 1:1.

\(^7\) Determining the appropriate attribute at which to split the tree at each branching is the key challenge in
constructing classification trees. Machine learning algorithms for creating tree classifiers generally make
the choice of the attribute on which to split so as to minimize entropy (Mitchell 1997, Quinlan 1986).
Entropy is defined per information theory (Shannon 1949), as essentially the degree of randomness or
uncertainty in the classification. However machine learning algorithms require large training data sets in
which the correct classification is known, and thus are more applicable in settings where an automated
system is an appropriate goal, as opposed to providing diagnostic feedback to improve human decision
making.
diagnosis of decision quality. The DDQ tree is used to classify each outcome event into diagnostic classes. In generic form, the diagnostic classes in a DDQ tree are as follows:

- **Favorable**: The performance outcomes were definitively favorable outcomes relative to a benchmark criterion.

- **Unfavorable**: The performance outcomes were definitively unfavorable relative to benchmark criterion. This can be further broken down into:
  - **Optimistic or Unlucky (OU)**: Unfavorable performance outcomes that can be explained by overly optimistic decision making, or by an unlucky turn of events (e.g., an unpredictable or random occurrence lead demand to be less than predicted).
  - **Pessimistic or Lucky (PL)**: Unfavorable performance outcomes that can be explained by overly pessimistic decision making, or by missed lucky opportunities (e.g., an unpredictable or random occurrence led demand to be more than predicted).
  - **Undetermined Problem**: Unfavorable outcomes that cannot be classified as OU or PL, at least without additional data.\(^8\)

- **Indeterminate**: It is not possible to diagnostically classify decision quality without further data or analysis.

In the DDQ methodology, the attribute on which to split the decision tree is determined at each node by the ability to make a definitive classification with a minimal increment in data and analysis requirements. This approach intuitively reflects both concerns for entropy reduction and the costs of data acquisition and computation of outcomes measures and criterion values. The construction of the DDQ Tree for hotel revenue management in essence is an extension of the previously discussed decision rule, expressed in terms of the three key variables: denials, average daily rate, and occupancy. Figure 2 shows the DDQ tree for the hotel revenue management context.

---

\(^8\) For example an outcome may not be clearly classifiable as OU or PL if it is a product of series of decision actions over time, some of which may have been optimistic and some of which may have been pessimistic.
Figure 2 – DDQ Tree for Hotel Revenue Management

The dotted lines on the DDQ tree show three levels of granularity of analysis. 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.** If there are no denials (Denials=0) for a given reservation night then, with the data we have at hand, the decision quality is indeterminate – no prospective customers were turned away so any errors in decision making did not affect outcomes. Analysis of decision quality under no denials would require further data about intermediary steps in the decision process (e.g., the accuracy of demand forecasts made – a key judgment). If Denials >0, then the key questions are: Were the denials necessary? and Were they made optimally to increase revenue?

**Occupancy.** If the hotel is not full, i.e., Occupancy < 100%, then, given that customers were turned away (Denials > 0), clearly the outcome is unfavorable, but to more completely categorize the outcome requires further analysis. If full occupancy is obtained (Occupancy = 100%) then denials are warranted. However, it is necessary to determine if they were made
correctly, that is, were low paying customers turned away so that high paying customers could be accommodated?

*Unconstrained Demand versus Capacity.* If Unconstrained Demand > Capacity, i.e., the ex-post estimate of unconstrained demand is more than hotel capacity, the situation is one of excess demand and we need to do further analysis at the next level of the tree. If Unconstrained Demand < Capacity, then since we have less than full occupancy and a positive number of denials, the outcome is classified as *Unfavorable – OU (Optimistic/Unlucky).* The latter 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 variance in demand led to an unlucky outcome, but the forecasting and pricing decisions were appropriate for the predictable component of demand. If there is excess demand, then some denials were warranted, but the lack of full occupancy indicates that some of the denials were not warranted.

\[ ADR_{\text{actual}} \] versus \[ ADR_{\text{market}} \]  Ceteris Paribus, 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 full occupancy is obtained. To assess improvements in ADR we compare \[ ADR_{\text{actual}} \] and \[ ADR_{\text{market}} \]. \[ ADR_{\text{actual}} \] is the average daily rate that the hotel actually obtained for the given reservation night. \[ ADR_{\text{market}} \] is estimate of the ADR that would be obtained if the sales mix of the hotel matched that the mix of customers in the market demand curve.\(^9\)

Where there is a full occupancy (the right side of the DDQ tree) and \[ ADR_{\text{actual}} \] exceeds \[ ADR_{\text{market}} \] then the outcome is classified as *Favorable*, otherwise the outcome is classified as *Unfavorable – PL (Pessimistic/Lucky).* 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 the random variance in demand led to a lucky outcome, but the forecasting and pricing was appropriate for the predictable component of demand.

\(^9\)\[ ADR_{\text{market}} \] may be estimated from data about days for which there are no denials. On zero denials days, the sales mix of the hotel is the same as the market mix. We assume that the proportion of willing customer across the range of price points is largely independent of the total unconstrained demand. Consequently, we can observe the market mix proportions on zero denial days and determine an \[ ADR_{\text{market}} \] value which we generalize to other reservation nights with non-zero denials within a given season and for a given day of the week.
On the left side of the DDQ tree, where there is less than full occupancy, but excess demand, a situation in which $\text{ADR}_{\text{actual}} > \text{ADR}_{\text{market}}$ indicates overly optimistic forecasting, too aggressive pricing or an unlucky random fluctuation in demand resulting in an outcome classification of Unfavorable-OU.

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 indeterminate diagnostic class.

Analysis of the resulting classifications provides evidence of the extent to which fluctuations in performance are due to systematic effects (such as decision making biases) as opposed to purely random variation. The distribution of outcomes classified as PL versus OU 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 tree provide information akin to the $G$ measure of the Lens Model, to the extent 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 to for further analysis of the DDQ tree classifications in the Diagnosis Phase.

Figure 3 shows the DDQ trees for “Big City” and “Southern Country”. Immediately we see the value of the DDQ approach. Both trees clearly show a majority of outcomes falling into the Unfavorable-OU category. Clearly, this suggests a systematic bias in the decision making towards overly optimistic forecasting and/or overly aggressive pricing. One immediate explanation is that this is not a bias in judgment at all but rather a function of the economic incentives of the task. In the hotel context, an OU 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, consequently, the economic incentive for the hotel is to be pessimistic in forecasting and conservative in pricing. Nevertheless, the RMs were too optimistic, as we explain further below.
Figure 3 – DDQ Trees for Alpha Hotels (1997-1999)
The Diagnosis Phase

To discover the context specific basis for the systematic patterns evidenced in the two DDQ trees for Alpha we move to the Diagnosis phase of our approach. Step 1 in the diagnosis phase, the initial identification of potentially relevant cues, is relatively simple when conducted with consultation from domain decision makers. 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 (Dawes 1979, Meehl 1954). In any event the diagnosis phase can be viewed as iterative – the analysis of DDQ charts sparks investigation of additional cues and the process of discovery of task information is thus initiated.

Step 2, the construction of DDQ control charts, entails creating a mapping between the DDQ tree diagnostic classifications and the relevant cues. Specifically, the DDQ Charts plot for each cue value (or range of cue values if the cue is continuous) the proportion of decisions classified into each diagnostic class. This can be at the macro level of Favorable, Indeterminate, Unfavorable or with the latter category further broken down into the OU and PL 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 values 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 as to systematic errors in decision making processes. Since some decisions are classified as indeterminate the favorable and unfavorable DDQ charts will not be simple reflections of each other and so there may be diagnostic value in plotting both. Charts at the more micro-level OU and PL classifications provide additional 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 OU class we plot only macro level DDQ charts.

In determining relevant cues over which to plot DDQ charts, several candidates appeared obvious, even before discussing them with domain experts: time, unconstrained demand, revenue manager experience, source of business. Time was clearly the prime candidate – demand for hotel rooms varies from season to season and within days of the week – and consequently the first control charts we constructed were looking for day of the week and seasonal fluctuations in the
systematic bias. Figure 4 shows the DDQ charts by day of the week for the two Alpha hotels. While the DDQ Tree showed a common systematic variance in outcomes, the DDQ charts suggest that the causes may be different at our two sites.

![Southern Country Unfavorable DDQ Classifications by Day of Week](image1)

![Big City Unfavorable DDQ Classifications by Day of Week](image2)

**Figure 4 DDQ Control Charts for Alpha Hotels**

We presented the DDQ trees and charts to management at Alpha in order to prompt the discovery of task information. The DDQ chart for Big City clearly shows that the RM consistently makes more systematic errors on Friday and Sunday nights. 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 confirmed the RM’s interpretation. The DDQ analysis thus led to the discovery that difficulties in estimating cancellations correctly were a primary cause of systematic failures in the RM’s decision making. Consequently Big City RM’s recognized the need to expend greater effort in estimating cancellations.

The DDQ chart for Southern Country shows that the RM at that location consistently makes more systematic errors on Wednesday and Thursday nights. The director of revenue management for the Alpha chain’s interpretation of this finding is worth noting because it was actually incorrect – demonstrating that the DDQ approach leads to hypotheses that then must be ratified by the data. The director suggested the RM used an aggressive pricing strategy throughout the week when it was only appropriate in the first part of the week. The director 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. The director’s hypothesis however was not borne out by the data – indeed there appeared to be no significant difference in the volume of group business between earlier and latter parts of the week. However, an analysis of demand at the highest price point showed
greater variability in demand on Wednesdays and Thursdays as compared with the other days of the week. Consequently we recommended Southern Country focus its efforts on improving forecasting of demand in the highest rate tier.

The DDQ chart provides substantial diagnostic information above that which could be obtained by simply plotting simple outcome variables over relevant cue values and attempting to observe systematic patterns. Since outcome variables are the product of both systematic and random variation, 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. The RevPAR analysis suggests problems on weekends at Southern Country, a problem attributable to the market environment in which Southern Country operates rather than a result reflecting any systematic failure in the quality of the revenue management decision making processes.

![Southern Country RevPAR by Day of Week](image)

Figure 5 – RevPAR by Day of Week for Southern Country

### 3.3 Discussion

The DDQ 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 economic incentives. We consider three possible explanations for the observed bias, from three different theoretical perspectives: behavioral decision theory, organizational behavior, and economics.
From a behavioral decision theory perspective, the bias can be interpreted as a failure to recognize “doing nothing” – inaction – as a decision alternative. Specifically, the RM closes price buckets (takes action) when they would have been better off doing nothing – leaving all price buckets open. Prior research suggests that, psychologically, action and inaction are quite different, yet normatively they should be evaluated on the same basis. For example, Kahneman and Tversky (1982) find greater regret occurs when unfavorable outcomes result from action rather than inaction (see also Landman 1987; Feldman, Miyamoto, & Loftus 1999; Zeelenberg, Van Der Pligt & de Vries 2000). Less regret from inaction is consistent with inaction not being recognized as something a decision maker “chooses” to do. Interestingly, the regret findings suggest a preference for inaction (see also Schweitzer 1994, Samuelson & Zeckhauser 1988), which appears to conflict with our findings at Alpha. However, if we accept that the basis of the regret difference lies in the perception that inaction is not “chosen”, that inaction is not viewed as a decision alternative, the apparent conflict can be resolved. Since inaction is something a decision maker is less responsible for than action, then the RM who obtains favorable outcomes from inaction is less likely to be rewarded by a supervisor than one who obtains similarly favorable outcomes from action. Since, psychologically, inaction is different from action, any 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).

The foregoing suggests that understanding behavior involves understanding the organizational context in which it takes places as well as the psychology of individual decision making. In such an organizational behavior context, issues such as roles, attributions and motivations come into play. Much of day-to-day behavior results from social roles (Fiske & Taylor 1991), and for the RM their day-to-day role is based on the need to monitor and take actions to close off price buckets. If price buckets are only rarely closed then there is little need for engaging a full time revenue manager to monitor reservations and make that decision. Furthermore, the pressure from senior management is always to improve revenues and the only short-term influence a revenue manager has over revenues is to be aggressive in pricing. Closing price buckets provides the possibility of revenue gains (and losses). Leaving price buckets open provides no opportunity for revenue gain other than by luck, it may however result in missed opportunities for revenue gains.
From an economic perspective we earlier noted that the price structure at Alpha (specifically the difference between the lowest and highest price points, relative to the lowest price point) creates an incentive for conservative behavior – for low balling forecasts and erring on the side of leaving price buckets open too long rather than closing them off too soon (a bias towards inaction). We attempted to quantify the actual monetary losses due to the opportunistic/aggressive bias. Specifically, by examining the price structure and the market mix it is possible to estimate the average revenue that would have been obtained from each customer that was denied a room erroneously (i.e., the left side of the DDQ tree – denials when there is less than full capacity). For Southern Country, the lost revenue amount to approximately 5% of the daily revenue that actually was obtained. This is equivalent to $250,000/year on a hotel with annual revenues of $5 million, more than enough to justify an additional full-time revenue manager or a substantial investment in a revenue management decision support system. The fact that some vendors of revenue management decision support systems now offer guarantees of 4% increase in revenues on adoption of their systems, suggests that this problem may not be unique to Alpha (Davern & Kauffman 2000). Regrettably, even with this information in hand, Alpha management have been slow in moving to implement even the most rudimentary decision support tools for revenue management.

4. Implications and Conclusions:

From a practical perspective, we have developed a methodology for providing diagnostic feedback in the field. As a field methodology, the DDQ approach specifically takes account of the costs of data acquisition and analysis. Such an approach is essential for organizations such as Alpha, where often they cannot afford to collect all the data necessary to easily assess the opportunity for improved decision making that appropriate investments in decision technologies and business processes could yield.

From a research methods perspective, our work on the DDQ methodology makes a key contribution to the literature. We provide an analysis of the defining characteristics of laboratory based judgment analysis methods. More importantly, we use this analysis to move from the lab to the field. Psychologically grounded field studies of decision making are substantially underrepresented in the literature, especially when compared to the plethora of laboratory studies

---

10 Non-disclosure agreements prevent us from providing actual dollar figures for Southern Country.
(there are a few notable exceptions, e.g., Melone 1994). Our work thus attempts to redress this imbalance. Of course, in so doing we have sacrificed internal validity in favor of richness and ecological validity, which places a caveat on the generalization of our findings.

This paper also addresses another imbalance in the literature – namely the under-representation of prescriptive decision making research, relative to normative or descriptive research (Bell, Raiffa & Tversky 1988). 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., Browne, Curley & Benson 1997).

From a theoretical perspective, we find field evidence that supports the laboratory findings that inaction and action are psychologically different, contrary to the normative perspective. Our study also serves to highlight the importance of understanding the context in which decision making takes place. Pressures from senior management and the definition of one’s role in an organization, clearly impact decision making processes – effects difficult to study in a laboratory environment.

More generally, we view this research as evidence of the value, both to practice and to science of drawing stronger connections between lab research and field research in decision making. In this vein, this paper illustrates the value of research building bridges between the richness of the field based study, and the internal validity of the lab study – with a view to providing useful prescriptions for practice, in addition to theory building and testing descriptive analyses.
References:


