December 2003

The Effect of Base Rate Sensitization on End-User Query Performance Moderated by Conscientiousness

A. Faye Borthick
Georgia State University

Paul Bowen
University of Queensland

David Robb
University of Queensland

Follow this and additional works at: http://aisel.aisnet.org/icis2003

Recommended Citation
Borthick, A. Faye; Bowen, Paul; and Robb, David, "The Effect of Base Rate Sensitization on End-User Query Performance Moderated by Conscientiousness" (2003). ICIS 2003 Proceedings. 65.
http://aisel.aisnet.org/icis2003/65

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2003 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
THE EFFECT OF BASE RATE SENSITIZATION ON END-USER QUERY PERFORMANCE MODERATED BY CONSCIENTIOUSNESS

A. Faye Borthick
Teaching and Learning with Technology Center
Georgia State University
Atlanta, GA USA
borthick@gsu.edu

Paul L. Bowen
School of Business
University of Queensland
Brisbane Australia
p.bowen@business.uq.edu.au

David A. Robb
School of Business
University of Queensland
Queensland, Australia
a.robb@business.uq.edu.au

Abstract

End users with extensive experience with an organization’s data can often detect query errors when query results do not correspond to their ex ante expectations. Many end users, for example, newly hired business analysts, however, compose queries on unfamiliar data. Their lack of familiarity means that they may be less able to evaluate the reasonableness of their query results. Although additional query experience will eventually give them the familiarity with the data that they need, in the interim, they may not recognize incorrect results from flawed queries. This paper develops and tests base rate sensitization as a means of enabling end users to improve their query performance. Contrary to the hypotheses, sensitizing end users to base rates, as a means of improving their assessments of the likely correctness of their query results, was not associated with significantly fewer query errors on a consistent basis. In a post hoc analysis, participant conscientiousness was found to moderate query performance. Participants of high conscientiousness that were sensitized to base rates made fewer query errors than those not sensitized. In contrast, base rate-sensitized participants with low conscientiousness made more errors than those not sensitized. In this interaction, high conscientiousness participants were able to take advantage of base rate information while low conscientiousness participants appeared to be hindered by base rate sensitization.

Keywords: Conscientiousness, end-user querying; query errors; base rate information

Introduction

After years of receiving and analyzing reports about their organizations, experienced managers and business analysts can classify query results as either reasonable or suspect. When confronted with suspect results, they can investigate whether business conditions have changed or whether the query that retrieved the data contained errors. From a Bayesian perspective, experienced
managers and analysts have developed internal base rate knowledge\(^1\) that they use to evaluate the reasonableness of query results (Biros et al. 2002; Klein et al. 1997; Roy and Lerch 1996). They can perform these evaluations before relying on the query results in their own work or making the results available to others in their organizations. Managers and business analysts that are not familiar with organizational data may not have developed the base rate knowledge necessary to evaluate their query results effectively (Ballou and Tayi 1999).

Organizations are increasingly relying on databases to support operations and provide timely information for decision making. Few organizations have enough experienced managers and business analysts, however, to satisfy the demands for query-proficient staff (Leonard-Barton 1995). Newly acquired managers and business analysts are often not sufficiently adept at classifying their query results as reasonable or suspect, a tendency that is exacerbated by end users’ tendencies to be overconfident in the correctness of their queries (Borthick et al. 2001).

Time and experience are prerequisites for evaluating query results correctly. Expediting this learning process for new managers and business analysts has the potential to reduce the probability of decision errors resulting from reliance on incorrect query results. This research investigates whether sensitizing new end users to using base rate information improves their query accuracy and their ability to align their confidence with the accuracy of their query results.

**Base Rate Sensitization and Query Performance**

**Signal Detection**

According to signal detection theory, an individual’s ability to discern between signal and noise is a function of the relationship between the response criterion and the discriminability of the distributions (Roy and Lerch, 1997). Discriminability decreases when the distributions overlap and the response criterion occurs in the overlap, increasing the difficulty of distinguishing between signal and noise. Figure 1 illustrates partially overlapped data distributions without and with errors.

![Figure 1. Overlapped Data Distributions Without and with Errors](image)


New managers and business analysts, who are likely to be unfamiliar with organizational data, confront a signal versus noise dilemma when evaluating query results. Distinguishing between signal and noise is especially difficult when the observed value occurs in the intersection of the two distributions (Figure 1). Distinguishing between signal and noise is further complicated when response criterion guidance is absent. Incorrectly interpreting a signal as noise is likely to result in decisions based on incorrect information. Incorrectly interpreting noise as a signal may result in additional time for reformulating queries and may lead to incorrect queries, giving erroneous results.

Individuals acting on instructions to heed the base rate frequency of signals have been found to improve their signal detection performance (Davies and Parasuraman 1981). Heeding base rate information has improved placement of the response criterion, thus providing the means for end users to discriminate between signal and noise (Roy and Lerch 1997). In the context of query formulation, providing base rate information allows end users to establish an anchor point for comparing with their query results.

---

\(^1\) A base rate refers to the relative frequency with which an event occurs or an attribute appears in a population (Ginossar and Trope 1987, Hinsz et al. 1988, Lanning 1987).
Significant divergence from the expected results should engender sufficient dissonance within the end user to signal a potential error in a query.

**Attention to Base Rates**

Kahneman and Tversky (1973) observe that decision makers appear to place insufficient weight on base rate information, to the detriment of decision quality. Research across several fields\(^2\) implies that decision makers should use base rate information to a greater extent. Base rates need not be highly reliable nor extreme to have diagnostic value (Koehler 1996).

In the absence of cues to the contrary (e.g., syntax error messages), inexperienced end users are apt to think they have prepared correct queries, leading to overconfidence in the quality of their decisions (Borthick et al. 2001). Being aware of base rate information, however, may create a standard of reference that can help them evaluate whether the information they extracted is reasonable (Figure 2). End users that are more sensitive to base rates, through experience or instruction, should be more likely to compare their results with historical base rates.

In analyses of management reports, one of the primary considerations is how current results compare with results of prior periods adjusted for seasonal differences. In the absence of major changes in the organization and its operating environment, analysts tend to expect current results to be similar to those of prior periods. If the results of two consecutive periods differ substantially, experienced analysts would search for explanations such as unintentional errors, timing problems, or deliberate misstatements. Accordingly, providing end users with historical base rate information in the form of management reports may have the potential to prompt end users to detect errors in their queries.

**Sensitization to Base Rates**

Constructing a contingency table and being sensitized to base rates through a graphical representation of the setting have been associated with making improved probability assessments (Roy and Lerch 1997). While the graphical representation had a greater positive effect, both sensitized groups answered more questions correctly than did control groups. Similarly, a combination of training, error incidence identification, and warnings about compromised data was more effective than traditional training for helping inexperienced end users detect compromised data (Biros et al. 2002).

Left to their own devices, inexperienced end users tend to believe they have formulated queries correctly once they pass syntax checks (Borthick et al. 2001). Experienced end users, however, realize that queries passing syntax checks may still be incorrect. Because of their familiarity with organizational data embodied in base rates, they have the capability of performing reasonableness checks on the query results.

If base rates were available to them and they learned to use them, inexperienced end users might be able to perform more like experienced end users in recognizing potential discrepancies between their query results and the base rate information. Once they recognize potential discrepancies, inexperienced end users might be able to reconsider and refine their query formulations. This reasoning yields the following hypothesis:

**H1:** Increasing base rate sensitization will be associated with decreasing query errors.

**Confidence in Query Correctness**

In the Brunswik Lens Model of human judgement (Brunswik 1956), individuals make judgments or predictions based on cues they sense from the environment that might be related to the criterion value they are attempting to discern or predict. In a database query setting, the criterion value corresponds to results from correctly formulated queries. Individuals’ judgements correspond to their queries, and the base rate information represents the cue set.

---

\(^{2}\)In information systems, Biros et al. (2002), Klein et al. (1997), and Roy and Lerch (1996); in accounting, Johnson et al. (2001) and Koonce (1993); in forensic science, Rogers, (2000); and in law, Cunningham and Reidy (1998).
If participants interpret base rate information (i.e., the cue set) correctly, the comparison is likely to prompt thoughtful end users to reflect on the correctness of their queries. Reflection by end users with higher levels of sensitization to base rate information on query results that do not match the base rate information is likely to induce greater dissonance. Even if end users make some adjustment to allow for a result that does not match the base rate information exactly, their adjustments are usually insufficient. These insufficient adjustments result in probability distributions that are too tight, i.e., they underestimate the range of actual occurrences (Tversky and Kahneman 1974). Accordingly, even for correct queries, this dissonance is likely to reduce the confidence that end users would otherwise have in the correctness of their queries. Furthermore, when query results differ substantially from the base rate information due to incorrect queries, this dissonance leads to end users being even less confident in the correctness of their queries. As a hypothesis, this prediction about the relationship between awareness and use of base rate information and end-user confidence is:

**H2:** Greater sensitization to base rate information will be associated with decreasing confidence in query correctness as the distance between the query result and base rate increases.

The Experiment

A laboratory experiment was conducted to test the hypotheses in a one-factor (presence or absence of sensitization to base rate information) with two covariates (complexity of the information request and grade point average (GPA)) between-groups experimental design.
Participants

Participants, 78 advanced undergraduate and postgraduate IS students, had previously learned to use the SQL query language and had practiced using SQL to develop queries. Participants were stratified according to their prior information systems and query experience and GPA and assigned to two groups, the top participant to Group A, the next participant to Group B, etc. The resulting equivalent groups were then randomly assigned to the control or treatment conditions.

<table>
<thead>
<tr>
<th>Table 1. Samples of Management Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management Report: Merchandise Line Mark Up Percentages July-August 2002</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Chemicals</td>
</tr>
<tr>
<td>Electrical</td>
</tr>
<tr>
<td>Audited Total Inventory Value at 1 Oct. Stock take</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Management Report: Sales and Deliveries July-August 2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Akkerman</td>
</tr>
<tr>
<td>Gillespie</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Total Gross Sales</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Less sales awaiting credit clearance</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Management Report: Supplier Deliveries July-August 2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>AEG</td>
</tr>
<tr>
<td>Apex</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Total gross purchases</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Percentage of Purchases delivered after want date since 1 July 2002</td>
</tr>
<tr>
<td>Percentage of total purchase rejects by supplier</td>
</tr>
</tbody>
</table>
Procedure

During the first hour (part 1) of the experiment, participants satisfied a sequence of information requests by preparing and executing SQL queries. A UNIX script captured each query attempt in text files, including start and end times for each request. All participants (treatment and control groups) received base rate information in the form of management reports. This report, illustrated in Table 1, was based on data in the database that participants queried. The treatment group was specifically directed to evaluate the correctness of query results based on related information for a prior period in the management reports. The control group received the same information requests and management reports but without the directions to heed the base rate information. During the second hour (part 2) of the experiment, participants prepared queries for similar information requests. Both groups received base rate information in the form of management reports. Neither the treatment nor the control group, however, received explicit directions to heed the base rate information. Table 2 shows a sample information request and its base rate sensitization for the treatment group.

After each query attempt was executed, the interface displayed the SQL result, i.e., either the syntax errors or the records produced by the query. Participants could revise their queries as many times as they wished until they were satisfied with their results. After responding to a prompt for their confidence levels in the correctness of their queries, participants began work on the next information request. Participants were not allowed to return to previously completed queries.

After eliminating incomplete responses, two examiners independently corrected participants’ responses. Each discrete alteration (addition or deletion of a query component) counted as one error. The examiners compared their independent assessments to ensure that all errors had been found and that the revised query formulations produced the correct results.

<table>
<thead>
<tr>
<th>Information request</th>
<th>For each employee, list their staff number, name, and the total retail dollar value of sales orders placed after 1 September 2002.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base rate sensitization for output</td>
<td>To maximize the likelihood of appropriately evaluating the correctness of your query output, establish a base rate by examining employee sales for the past quarter. The treatment group management report contains values for the last quarter (3 months). Your required dollar value is for one month (1 September to 1 October). Therefore total sales on your query output should be approximately 1/3 of the total sales for the quarter. Use the base rate to evaluate the correctness of the output of your query.</td>
</tr>
</tbody>
</table>

Analysis of Results

Planned Analyses

Preliminary analysis of the part 1 queries for hypothesis 1 indicated no statistically significant results between the control and treatment groups with respect to query performance. The least-squares means for the number of errors for the control (8.5724) and treatment groups (9.2602) during part 1 of the experiment were not significantly different. Furthermore, the differences were opposite to the hypothesised direction. For part 2, the control and treatment groups’ least-squares means for the number of errors were 7.4207 and 8.3706, respectively. Again, not only was there no significant difference between the performance of the control and treatment groups, but the differences were not in the hypothesised direction.

Post Hoc Analysis

Given the unexpected results, previously collected participant personality data were included in the statistical analysis to investigate the potential moderating effect of participant conscientiousness on performance. With conscientiousness (Costa and McCrae 1992) as a covariate, there were statistically significant differences in performance of the treatment and control groups (Figure 3). Treatment group participants with above average levels of conscientiousness made significantly fewer errors than their counterparts in the control group. Although the difference was not statistically significant, treatment group participants with high
conscientiousness also made fewer errors than their control group counterparts. In contrast, treatment group participants with low and below average levels of conscientiousness made significantly more errors than their control group counterparts. These results indicate the presence of an interaction effect between conscientiousness and the usefulness of base rate information. That is, conscientious participants were able to take advantage of the base rate information to improve their performance. In contrast, the presence base rate information was associated with impaired performance for participants with lower levels of conscientiousness.

![Figure 3. Performance Moderated by Conscientiousness](image_url)

**Implications**

If sensitization to base rates improved query performance for everyone, organizations might be well served by incorporating base rate sensitization into training and mentoring programs for new managers and business analysts. But contrary to a one-size-fits-all model, post hoc analysis of participant queries detected an interaction effect. That is, while base rate-sensitization of participants with above average conscientiousness was associated with fewer query errors, sensitized participants with below average conscientiousness made more query errors. These results imply that if organizations consistently hire conscientious employees, sensitization to base rate information is likely to be beneficial. Organizations that hire less conscientious employees may need to place greater emphasis on helping users learn to formulate and evaluate queries systematically. Whether organizations hire employees with low or high conscientiousness, promoting mentoring relationships between experienced and new managers and business analysts is likely to improve end users’ query performance.

**References**


