



PERGAMON

Omega 29 (2001) 391–404

omega

The International Journal  
of Management Science

www.elsevier.com/locate/dsw

# Detecting errors in data: clarification of the impact of base rate expectations and incentives

Barbara D. Klein\*

*School of Management, University of Michigan-Dearborn 4901 Evergreen Road, Dearborn, MI 48128-1491, USA*

Received 17 April 2000; accepted 31 March 2001

## Abstract

Organizational databases have a significant rate of data errors and detecting and correcting these errors can be problematic. This paper builds on a stream of research demonstrating that users of these databases can detect data errors under certain circumstances. A theory of error detection and research on the effect of base rate expectations in probabilistic judgement tasks are applied to the development of two propositions about error detection. It is argued that expectations about the base rate of errors in data affect error detection performance when they are developed through direct experience and that incentives affect error detection performance. The two research propositions are tested in a laboratory experiment. Experience-based expectations about the base rate of errors and incentives are found to affect error detection performance. © 2001 Elsevier Science Ltd. All rights reserved.

*Keywords:* End-user computing; Data quality; Error detection

## 1. Introduction

Between one and 10% of data items in critical organizational databases are estimated to be inaccurate [1,2], suggesting a large potential problem if these data are used for managerial decision making. Though some (e.g. [3]) have argued for system-based error detection, practical limitations of this approach to improving data quality suggest that system-based error detection alone will not provide an adequate solution in most organizations.

System-based error detection is limited by the following factors. First, organizations may not be willing to pay the system development costs associated with preventing data errors. Second, even when organizations are willing to pay the costs associated with system-based error detection, it is difficult to predict all data errors that might occur. Third, even if system-based approaches ensure high-quality data at one point in time, over time data may become inaccurate if the entities and attributes described by the data change while the corresponding data do not change. Orr [4] notes

that data that are not used by humans will soon be incorrect. Fourth, users may have different understandings about the meaning of certain data fields. System-based approaches cannot solve this kind of problem. Fifth, even if systems developed by MIS professionals contain reasonably effective error detection algorithms, systems developed by end users are much less likely to contain such programs.

Given these limitations, system-based error detection alone will not typically result in high quality data. Indeed, Orr notes that “no serious information system has data quality of 100%” [4]. For most organizations, the optimal approach to error detection should probably rely on both human and system-based approaches to error detection [5].

Several studies have been critical of human ability to find data errors [1,6,7]. However, recent evidence collected in both laboratory experiments [5] and in field studies [8] demonstrates that users of information systems can and do detect data errors in certain circumstances and that human error detection performance can be improved. In Klein et al. [8], evidence from interviews with professionals in three domains suggests that high levels of incentives to detect data errors, perceptions about the materiality of data errors, and perceptions about the base rate of errors are associated with

\* Tel.: +1-313-593-5268; fax: +1-313-593-5636.

E-mail address: bdklein@umd.umich.edu (B.D. Klein).

incidents of error detection. Klein et al. [5] used a complementary pair of laboratory experiments to test these contentions in a controlled fashion. They found that user's error detection performance is influenced by explicit goals and incentives in expected ways, but they found no significant effect of base rate expectations.

One explanation of why Klein et al. [5] did not find a relationship between base rate expectations and error detection performance is that base rate expectations were manipulated by giving subjects information about the historical error rate of the information with which they worked. In contrast to the information-based expectations of subjects in the laboratory experiment, the base rate expectations of the professionals interviewed in the field study had been developed through several years of direct experience working with data. The conclusion drawn from the disparate findings of the laboratory and field studies is that users will respond to the experience of detecting errors in data rather than to information about the base rate of errors. It appears that expectations about the error rate are most strongly influenced by hands-on experience with error detection.

This conclusion is consistent with current thinking on the effect of base rate expectations in probabilistic judgement tasks. Probabilistic judgements have been found to be affected by base rate expectations when these expectations are developed through direct experience [9].

This paper presents the findings of a laboratory experiment testing the effect of experience-based expectations about the base rate of errors on error detection performance. To date, the only evidence supporting this relationship has been found in a field interview study that relies on self-reported incidents of error detection. Evidence collected in a controlled laboratory setting would clearly increase our confidence in the existence of a relationship between experience-based expectations and performance. The laboratory experiment discussed in this paper also provides further evidence about the relationship between incentives and error detection performance. While the main emphasis is on understanding the way in which base rate expectations affect performance, the evidence on incentives and performance provides a partial replication and an extension to the work discussed in Klein et al. [5].

The remaining sections of this paper present (1) a review of prior research on the question of user detection of errors in data, (2) a theory of error detection, (3) the research propositions and experimental methodology, (4) the results of a laboratory experiment, (5) a discussion of the experimental results, and (6) conclusions.

## 2. Background

Researchers working in the area of data quality agree that data accuracy, currency, and completeness are important areas of concern (e.g. [10–13]). This study is primarily concerned with data accuracy which has been defined as

agreement with an attribute of a real world entity, a value stored in another database, or the results of an arithmetic computation [12].

A number of broad conclusions can be drawn from the research on data quality. First, error rates significantly greater than zero have been found in all of the studies that have measured the rate of errors in data (e.g. [1,14,15]). Second, while some researchers propose methods designed to completely rid databases of errors (e.g. [16]), others propose tools for determining how to best allocate limited resources to controlling the level of data errors [3,17–19]. Third, some researchers recognize that purging all errors from databases is sometimes an unrealistic objective and suggest a variety of approaches for using imperfect data [17,20–24].

Three early studies in the data quality literature present considerable evidence of poor user performance in detecting data errors. Davis et al. [6] conducted a field experiment in which individuals were mailed banking confirmation statements with embedded errors. Approximately half of the individuals failed to detect important errors in their account information. Despite a warning stating that users should verify the accuracy of information provided by the FBI, Laudon [1] found that users of criminal information systems rarely detect errors in information provided to police departments by the FBI. Ricketts [7] conducted a laboratory experiment in which over 90% of the subjects failed to detect a substantial data error in production planning reports. The failure of humans to detect errors in data is also an underlying assumption in much of the literature on data quality that advocates for the upfront improvement of data quality in organizational databases [2].

More recent studies show that users of information systems detect data errors under some conditions. In a study investigating whether actuaries typically detect errors in data, all 10 of the actuaries interviewed reported instances in which they detected errors in data. The degree to which data items were reviewed for errors prior to use was affected by the actuaries' expectations about the likelihood of errors in a given dataset. The actuaries did not attempt to detect all errors, apparently responding to a perceived tradeoff between the time and effort required to find additional errors versus the potential impact of a more accurate dataset [25].

Two laboratory experiments demonstrate that explicit goals and incentives improve users' error detection performance, but that giving users information about the error rate does not affect performance [5]. A field study shows that incentives, the materiality of data errors, and expectations about the error rate developed through direct experience improve error detection performance [8].

## 3. A theory of error detection

There are four possible outcomes in error detection tasks, as shown in Fig. 1 [26]. Data errors can be successfully detected (hits) or missed (misses). When no error exists, an

		Behavior	
		Error Detected	Error Not Detected
Data	Error Exists	Hit	Miss
	Error Does Not Exist	False Alarm	Correct Rejection

Fig. 1. Presence of errors and detector responses.

individual may incorrectly conclude there is an error (false alarm) or correctly conclude there is no error (correct rejection). The performance of an individual can be characterized by a table like the one in Fig. 1 with a conditional probability in each cell. Research on quality control inspection in manufacturing systems uses a similar conceptualization of error detection behaviour [27–29].

### 3.1. Signal detection theory and error detection

Signal detection theory is an extension of statistical decision theory that gives us a better understanding of how situational factors affect the likelihood of the four outcomes in Fig. 1. Signal detection theory has been applied to a broad range of tasks such as the detection of a tone [30], the detection of a flash of light [30], the detection of social cues [31], recognition memory [32], and human vigilance [33]. In all of these studies there is an interest in explaining the performance of humans distinguishing between two events. Signal detection theory labels these two events as “noise” and “signal”. Noise can be thought of as background, and a signal is a stimulus that may be detected because it deviates from background noise. Signal detection theory explains the performance of humans who are trying to determine when a signal is present [30,34–36].

When we apply signal detection theory to the detection of data errors, a record without errors represents background noise, and a record with a data error provides a signal. Typically, a user must examine a display of information and determine whether the data contained in the display are correct or errorful. Records with errors and records without errors present different aspects to the user, and this presumably affects the user’s perception of the record. Following signal detection theory, this can be represented by the two distributions in Fig. 2 [37].

The two curves shown in Fig. 2 represent two different distributions of the internal state of the human detector that mediates between a stimulus and a response [37]. One of the curves represents the distribution of the internal state that occurs in the presence of random noise. The other curve represents the distribution of the internal state that occurs when a signal is added to random noise. A record with a glaring error might be perceived by a user as being near the right-hand side of the figure, to the right of  $x_2$ . These the user could reliably classify as errorful. A record with no hint of problems might be perceived as being near the left-hand side of the figure, to the left of  $x_1$ . These the user could reliably

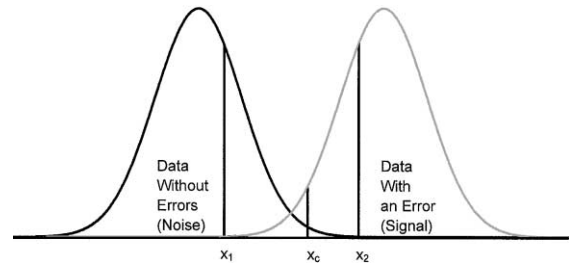


Fig. 2. Distributions for data with an error and data without errors.

classify as correct. In general, it is assumed that a user is not able to perfectly distinguish correct from errorful records in all cases [36,37]. Thus, the two distributions overlap. Records perceived as being in the overlapping portion of the figure (between  $x_1$  and  $x_2$ ) cannot be so easily classified.

The stimulus variable on the  $x$ -axis in Fig. 2 has a more abstract meaning when the diagram is applied to human-based error detection than it typically does in signal detection. In most signal detection research,  $x$  is the magnitude of some explicit signal such as the loudness of a tone or the brightness of a light. In human-based error detection,  $x$  is a “discomfort index” assumed to be generated by the human in the course of looking at the data.

Two factors drive the performance of individuals attempting to distinguish between signal and noise: the discriminability of the distributions and the response criterion. As the overlap between the two distributions increases, discriminability decreases and a greater portion of stimuli cannot be reliably classified. With decreased discriminability, false alarms or misses become more frequent. High discriminability would mean that if there were an error, it would tend to be very glaring. Thus it would be easy to discriminate between correct and errorful records. Low discriminability would mean that many records had ambiguous clues that might or might not reflect actual errors.

Signal detection theory suggests that human detectors set a response criterion at some point between  $x_1$  and  $x_2$  to determine whether to label an ambiguous stimulus as noise or as a signal. An example response criterion is shown in Fig. 2 as  $x_c$ . Perceptions to the right of  $x_c$  would be labeled signal, and perceptions to the left of  $x_c$  would be labeled noise. An individual might set his or her response criterion at any point between  $x_1$  and  $x_2$ . For example, a user might decide to ignore all but the most glaring clues about errors (i.e., place

the response criterion far to the right). Alternatively, a user could decide to flag any even slightly ambiguous record as errorful (i.e., place the response criterion far to the left).

Signal detection theory suggests that expectations about the base rate of errors will influence the placement of the response criterion. A user who expected that data from a particular source would have many errors might tend to set his or her response criterion rather far to the left. Contrariwise, if a user expected that most data from a second source was error free, he or she might set the response criterion rather far to the right [26].

Incentives derived from the costs and benefits of detecting or missing errors should also affect the placement of the response criterion. If a missed error is extremely damaging, the response criterion would be set far to the left. If a false alarm is extremely embarrassing, then the response criterion would be set far to the right [26].

In most applications of signal detection theory it is assumed that discriminability is fixed for any given task such that individuals can adjust their response criteria but they cannot adjust discriminability. This is a reasonable assumption in the strictly perceptual task domains (such as the auditory detection of a tone) for which signal detection theory was originally developed. When discriminability is fixed, individuals cannot increase the rate of hits without increasing the rate of false alarms. Consistent with this view of discriminability, signal detection theory does not generally address the link between effort and performance. However, in applying signal detection theory to the detection of errors in data, there may be changes in discriminability due to shifts in attention or effort in this semantic discrimination task. By trying harder or paying closer attention to a display of information, an individual may be better able to distinguish errors from correct data values. This suggests that effort may affect performance through changes in discriminability.

Signal detection theory thus provides a way of conceptualizing the error detection process as one in which individuals choose where to set their response criteria. They may also choose how much effort to expend in analyzing a data display, thus modifying their ability to discriminate between correct and errorful records. Changes in error detection performance can be analyzed in terms of shifts in discriminability and shifts in response criteria.

### 3.2. Borrowing from additional theory bases

Several other theories are useful in understanding error detection performance: theories of individual task performance, theories of effort and accuracy in decision making, theories of incentives, and theories of base rate expectations in probabilistic judgment tasks.

#### 3.2.1. Theories of individual task performance

Theories of individual task performance provide some insight into conditions under which users can improve their discriminability. Campbell's theory of individual task

performance suggests that effort may affect error detection performance, presumably by affecting an individual's ability to successfully discriminate between correct and errorful records [38].

#### 3.2.2. Theories of effort, accuracy, and incentives in decision making

A number of studies based on theories of effort and accuracy in decision making have found that incentive structures can influence the adoption of information processing strategies and hence task performance (e.g., [39]). Numerous other studies have demonstrated that performance-contingent incentives affect task performance in both laboratory and field settings (e.g. [40–42]). In the context of the error detection task, individuals may allocate their total effort (including effort focused on improving discriminability) so that they improve their total payoffs.

#### 3.2.3. Theories of base rate expectations in probabilistic judgment tasks

Koehler [9] recently synthesized the literature on base rate expectations in probabilistic judgment tasks. The synthesis recognizes that there are conditions under which people make less use of base rate information than prescribed by a Bayesian normative standard [43–45]. However, there are a number of task conditions that affect the extent to which base rate information is used [9]. Probabilistic judgments have been found to be affected by base rates under three conditions: (1) when base rate expectations are developed through direct experience [58,46–48], (2) when base rate information is based on multiple, random samples from clearly defined sample spaces [43,49,50], and (3) when base rates are based on large samples or provided by reliable sources [49,51]. More recently, Roy and Lerch [52] show that performance in probabilistic judgment tasks is improved when base rate information is presented graphically rather than in a verbal problem statement.

In this investigation, we are primarily interested in the finding that people make use of base rate expectations when these expectations are developed through direct experience, but not when they are simply given information about base rates.

## 4. Research propositions and methodology

The theory of error detection outlined above can be summarized in two propositions about the conditions under which individuals will detect errors in data. A laboratory experiment was conducted to test the research propositions. The research propositions, experimental task, experimental treatments, dependent variables, and the subjects and sample size are discussed in turn.

#### 4.1. Research propositions

**Proposition 1.** *Expectations about the base rate of errors in data developed through direct experience will influence performance in the detection of data errors.*

Specifically, there will be both more hits and better discrimination when expectations about the base rate of errors are higher. Discrimination refers to an individual's ability to distinguish data errors from correct values and is recognized by an increase in the hit rate without an accompanying increase in the false alarm rate.

**Proposition 2.** *Incentives will influence performance in the detection of data errors.*

Incentives to detect data errors will improve the hit rate and discrimination. Empirical evidence supporting this proposition was found in an earlier experiment [5].

#### 4.2. Experimental task

The experimental task was the calculation of accrued pension benefits for employees in several divisions of a firm, using data containing a number of embedded errors. After a brief review of the spreadsheet commands needed to manipulate data, subjects were given 110 min in which to perform the experimental task for at least two, but as many as possible, of six divisions. The time limit was imposed in order to simulate organizational settings in which users who have several tasks to complete during a work day have to decide how much time to allocate to error detection.

Subjects were given three types of materials to assist them in their task: (1) a data file for the current year in spreadsheet format, (2) printed reports with the current and the previous year's data for each division to facilitate comparisons over time, and (3) a procedure manual explaining the meaning of the data elements and the algorithm for the pension calculation. The procedure manual also mentions that errors are sometimes found in the data provided by client companies and that when errors are found a confirmation memo should be prepared to be sent to the client so that the client can confirm or disconfirm that the data items are in error. A short questionnaire containing items for manipulation checks was administered at the end of the experimental session.

Errors were embedded in 10 of the 50 employee records that subjects received for a target division. An example of an error in an employee record is an inconsistency between the value in the "hire date" field and the value in the "years of service" field. A domain expert verified that the types of errors embedded in the data are representative of the data errors that occur in the domain. The embedded errors for the target division were the same in all of the experimental conditions.

Table 1  
Experimental design (cells with a \$100 prize are noted)

Incentives	Base rate expectations		
	1. High	2. Control	3. Low
1. Detection	\$100 prize	\$100 prize	\$100 prize
2. Random	\$100 prize	\$100 prize	\$100 prize
3. No Incentive	No prize	No prize	No prize

#### 4.3. Experimental treatments

The experimental treatments divided subjects into three base rate expectations (High, Low, and Control) and three incentives (Detection, Random, and No Incentive).

##### 4.3.1. Expectations about the base rate of errors

The expectation construct was operationalized through direct experience. Subjects were not told how many errors were present in the datasets. Instead, in the High base rate expectations condition, subjects performed the pension calculations using a dataset containing an error rate of 30% (15 of 50 records contained an error) before performing the pension calculations for a second target division. On average, subjects in the High base rate expectations condition actually found 6.8 out of 15 errors in the data for the first division. In the Low base rate expectations condition, subjects performed the pension calculations using a dataset containing an error rate of 4% (2 of the 50 records contained an error) before performing the pension calculations for a second target division. On average, subjects in the Low base rate expectations condition actually found 0.09 out of 2 errors in the data for the first division. For the Control condition, subjects performed the pension calculations for the target division before doing the task for any other divisions. In the Control condition subjects had no prior experience from which to develop base rate expectations. The error rate for the target division was the same (10 of the 50 records contained an error) for all the experimental conditions.

##### 4.3.2. Incentives

The incentives construct was operationalized through \$100 prizes. A lottery procedure found to be effective in prior research (e.g. [5,39]) was used to award the prizes. Altogether six \$100 prizes were awarded (see Table 1).

In the Detection condition, subjects were told that their performance on the task would be evaluated and that the 30% of subjects with the best performance would be entered into a lottery from which one person would receive \$100. These subjects were told that performance on the confirmation memo would be weighted three times as heavily as performance on the accrued pension calculations. In the

Random condition, subjects were told that everyone who conscientiously participated in the experiment would be entered into a lottery from which one person would receive \$100. No prizes were awarded in the No Incentive condition.

Instructions about the prizes were given to subjects in the Detection and Random conditions at the beginning of the experiment and were displayed prominently on an overhead projector screen throughout the experiment. Subjects in these conditions had an a priori chance of 1 in 18 of winning the prize making these incentives salient to them.

#### 4.4. Dependent variables

Four dependent variables were used: the proportion of errors successfully detected for the target division (i.e., hits), the proportion of correct records misidentified as errors for the target division (i.e., false alarms), and two measures of discrimination. In signal detection theory, discrimination is usually based on several different trials for each subject. In this experiment there is only a single hit-rate/false-alarm-rate pair for each subject. We use a technique that estimates discrimination when only a single hit-rate/false-alarm-rate pair is available [30,53,54]. This measure will be called  $\text{Discrimination}_{\text{SD}}$  (based on Signal Detection Theory).  $\text{Discrimination}_{\text{SD}}$  varies from 0.5, which is random performance, to 1.0 which is perfect performance. See Appendix A for more detail.

Because  $\text{Discrimination}_{\text{SD}}$  is an estimate of discrimination, a second measure of discrimination from behavioral decision research is also used [55]. This measure conceptualizes discrimination as the extent to which items are correctly sorted. For example, a person asked to separate defective and perfect items exhibits perfect discrimination if all defective items are placed in one pile and all perfect items are placed in a second pile. This measure of discrimination, called  $\text{Discrimination}_{\text{BD}}$  (based on Behavioral Decision Theory) in this study, is a measure of the extent to which the proportion of defective items in each pile deviates from the base rate of defective items. In this study,  $\text{Discrimination}_{\text{BD}}$  varies from 0, which is random performance, to 0.16 which is perfect performance. See Appendix A for more detail.

#### 4.5. Subjects and sample size

Students from upper-level undergraduate and MBA business courses participated in the experiment. Students were used because they are less likely to have beliefs about the base rate of errors in data than experienced professionals.

Subjects were randomly assigned to the nine conditions (see Table 1). A power analysis suggested that a sample size of 18 observations per condition would be sufficient to detect an effect size equal to one standard deviation of the distribution of the dependent variables with a power of 0.95. This gave a total of 162 observations.

## 5. Results

Table 2 presents the results of a standard ANOVA for each of the four dependent variables. Main effects of the incentives factor on hits, false alarms, and  $\text{Discrimination}_{\text{BD}}$  were found. For the base rate expectations factor, main effects were found for all four of the dependent variables. Interaction effects between the incentives factor and the base rate expectations factor were found for false alarms,  $\text{Discrimination}_{\text{SD}}$  and  $\text{Discrimination}_{\text{BD}}$ . The power of the tests for which statistically significant effects were not found is greater than 0.98.

An analysis of the residuals was conducted to determine whether the ANOVA models are appropriate for the data collected in the experiment. When the hit rate and the two measures of discrimination are the dependent variables, no departures from the assumptions of independence of error terms, normality of error terms, and equality of population variances were detected. Additionally, no outliers were detected. When the false alarm rate is the dependent variable, the assumption of independence of error terms was not violated and no outliers were detected. However, the assumptions of normality of error terms and equality of population variances were violated [56]. Additional non-parametric tests were therefore conducted to test for the effect of base rate expectations and incentives on false alarms. These non-parametric tests are not affected by the violation of the assumptions of the ANOVA model.

An additional variable was added to the ANOVA models for each of the four dependent variables to test whether there were differences in performance between graduate and undergraduate students. No statistically significant differences were found.

A discussion of the effect of base rate expectations and incentives on each dependent variable follows. Treatment means and paired comparisons for each dependent variable are presented. Results of the non-parametric tests for the false alarm dependent variable are also discussed.

#### 5.1. Effect of base rate expectations and incentives on the hit rate

The treatment means showing the effect of the incentives and base rate expectation factors on hits are shown in Table 3.

The mean number of hits ranges from 2.4 to 7.5 across the experimental conditions. Since it was possible to detect 10 errors, the hit rate ranges from 0.24 to 0.75. The Tukey method of multiple comparisons was used to conduct pairwise comparisons of the treatment means for hits. Since there is no interaction effect between the incentives and base rate expectations factors for the hit dependent variable, the paired comparisons were performed in terms of the factor level means [56]. Using a family confidence coefficient of 0.95 for the incentives factor, it was concluded that subjects in the Detection condition had more hits than subjects in the

Table 2  
ANOVA results

	Hits	False alarms	Discrimination <sub>SD</sub>	Discrimination <sub>BD</sub>
Main effect of incentives	Yes ( <i>p</i> = 0.0001)	Yes ( <i>p</i> = 0.0001)	No	Yes ( <i>p</i> = 0.0001)
Main effect of base rate expectations	Yes ( <i>p</i> = 0.0001)	Yes ( <i>p</i> = 0.0001)	Yes ( <i>p</i> = 0.0001)	Yes ( <i>p</i> = 0.0001)
Interaction effect	No	Yes ( <i>p</i> = 0.0001)	Yes ( <i>p</i> = 0.0001)	Yes ( <i>p</i> = 0.0001)

Table 3  
Effects of base rate expectations and incentives on hits<sup>a</sup>

		Base rate expectations		
		High	Control	Low
Incentives	Detection	7.5 (0.75)	5.6 (0.56)	3.4 (0.34)
	Random	5.1 (0.51)	3.4 (0.34)	3.0 (0.30)
	No incentive	3.9 (0.39)	3.3 (0.33)	2.4 (0.24)

<sup>a</sup>Each cell gives the mean number of hits followed by the mean hit rate in parentheses.

Random condition and that subjects in the Detection condition had more hits than subjects in the No Incentive condition. The following conclusions were drawn for the base rate expectations factor using a family confidence coefficient of 0.95: (1) subjects in the High condition had more hits than subjects in the Control condition, (2) subjects in the High condition had more hits than subjects in the Low condition, and (3) subjects in the Control condition had more hits than subjects in the Low condition.

5.2. *Effect of base rate expectations and incentives on the false alarm rate*

The treatment means showing the effect of the incentives and base rate expectation factors on false alarms are shown in Table 4.

The mean number of false alarms ranges from 0.06 to 7.17 across the experimental conditions. Since it was possible to have 40 false alarms, the false alarm rate ranges from 0.002 to 0.179. With the exception of subjects assigned to the Detection level of the incentives factor and the Low level of the base rate expectations factor, subjects had on average fewer than one false alarm. The Tukey method of multiple comparisons was used to conduct pairwise comparisons of the treatment means for false alarms. Since there is an interaction effect between the incentives and base rate expectations factors that cannot be removed through a transformation, the paired comparisons were performed in terms of the treatment means [56]. Using a family confidence coefficient of 0.95, it was concluded that subjects assigned to the

Detection level of the incentives factor and the Low level of the base rate expectations factor had more false alarms than subjects in all of the other eight conditions. No other significant differences in the paired comparisons were found.

The findings of non-parametric tests conducted because of the violations of the assumptions of the ANOVA model tell a similar story. Table 5 shows the number and percentage of subjects at each level of the base rate expectations factor with one or fewer false alarms and with more than one false alarm. The conclusion from a test of the difference between these population proportions [57] is that the proportions are not equal. Thus, there is a relationship between base rate expectations and false alarms. Subjects with Low base rate expectations are more likely to have more than one false alarm than subjects in the control condition, and subjects with High base rate expectations are less likely to have more than one false alarm than subjects in the control condition.

Table 6 shows the number and percentage of subjects at each level of the incentives factor with one or fewer false alarms and with more than one false alarm. The conclusion from a test of the difference between these population proportions [57] is that the proportions are not equal. Thus, there is a relationship between incentives and false alarms. Subjects with incentives to detect data errors are more likely to have more than one false alarm than other subjects.

5.3. *Effect of base rate expectations and incentives on Discrimination<sub>SD</sub>*

The treatment means showing the effect of the incentives and base rate expectation factors on Discrimination<sub>SD</sub> are shown in Table 7.

The Discrimination<sub>SD</sub> scores shown in Table 7 range from 0.584 to 0.937. A score of 0.50 reflects a complete lack of discrimination (i.e., random performance), and a score of 1.0 reflects perfect discrimination. Discrimination<sub>SD</sub> scores greater than 0.80 are generally considered to reflect good discrimination. The Tukey method of multiple comparisons was used to conduct pairwise comparisons of the treatment means for Discrimination<sub>SD</sub>. Since there is an interaction effect between the incentives and base rate expectations factors that cannot be removed through a transformation, the

Table 4  
Effects of base rate expectations and incentives on false alarms<sup>a</sup>

		Base rate expectations		
		High	Control	Low
Incentives	Detection	0.06 (0.002)	0.44 (0.011)	7.17 (0.179)
	Random	0.06 (0.002)	0.22 (0.006)	0.67 (0.017)
	No incentive	0.28 (0.007)	0.50 (0.013)	0.56 (0.014)

<sup>a</sup>Each cell gives the mean number of false alarms followed by the mean false alarm rate in parentheses.

Table 5  
Number and percentage of subjects with more than one false alarm at each level of base rate expectations

Base rate expectations	False alarms ≤ 1	False alarms > 1
High	53 (98.1%)	1 (1.9%)
Control	48 (88.9%)	6 (11.1%)
Low	38 (70.4%)	16 (29.6%)

Table 6  
Number and percentage of subjects with more than one false alarm at each level of incentives

Incentives	False alarms ≤ 1	False alarms > 1
Detection	39 (72.2%)	15 (27.8%)
Random	51 (94.4%)	3 (5.6%)
No Incentive	49 (90.7%)	5 (9.3%)

Table 7  
Effects of base rate expectations and incentives on Discrimination<sub>SD</sub>

		Base rate expectations		
		High	Control	Low
Incentives	Detection	0.937	0.881	0.584
	Random	0.862	0.789	0.779
	No Incentive	0.815	0.724	0.751

paired comparisons were performed in terms of the treatment means [56]. Using a family confidence coefficient of 0.95, the following conclusions can be drawn: (1) for the Detection level of the incentives factor, subjects with High base rate expectations had better discrimination than subjects with Low base rate expectations and subjects in the Control condition had better discrimination than subjects in the Low condition, (2) subjects assigned to the Detection level of the incentives factor and the Low level of the base rate expectations factor had poorer discrimination than subjects assigned to all of the other conditions except the No Incentive and Control base rate expectations condition.

Table 8  
Effects of base rate expectations and incentives on Discrimination<sub>BD</sub>

		Base rate expectations		
		High	Control	Low
Incentives	Detection	0.1133	0.0769	0.0283
	Random	0.0730	0.0456	0.0373
	No Incentive	0.0540	0.0441	0.0285

5.4. Effect of base rate expectations and incentives on Discrimination<sub>BD</sub>

The treatment means showing the effect of the incentives and base rate expectation factors on Discrimination<sub>BD</sub> are shown in Table 8.

The Discrimination<sub>BD</sub> scores shown in Table 8 range from 0.0283 to 0.1133. The highest possible Discrimination<sub>BD</sub> in this experiment is 0.16. The Tukey method of multiple comparisons was used to conduct pairwise comparisons of the treatment means for Discrimination<sub>BD</sub>. Again, since there is an interaction effect between the incentives and base rate expectations factors that cannot be removed through a transformation, the paired comparisons were performed in terms of the treatment means [56]. Using a family confidence coefficient of 0.95, the following conclusions can be drawn: (1) subjects assigned to the Detection incentive and the High base rate expectations condition had better discrimination than subjects in all of the other conditions, (2) subjects assigned to the Detection incentive and the Control base rate expectations condition had better discrimination than subjects assigned to the three experimental conditions with Low base rate expectations, and (3) subjects assigned to the Random incentive and the High base rate expectations condition had better discrimination than subjects assigned to the three experimental conditions with Low base rate expectations.

6. Discussion of experimental results

The results relative to the two research propositions are summarized in Table 9. A discussion of the results for each proposition follows.

Table 9  
Summary of results for research propositions

Error detection performance will be influenced by:	Research results
1. Expectations about the base rate of errors	Main effect on the hit rate  Effect on the false alarm rate and discrimination through an interaction with incentives
2. Incentives	Main effect on the hit rate  Effect on the false alarm rate and discrimination through an interaction with base rate expectations

**Proposition 1.** *Expectations about the base rate of errors in data developed through direct experience will influence performance in the detection of data errors.*

There is strong evidence that base rate expectations had an impact on the hit rate. Subjects with High base rate expectations had more hits than subjects in the control group, and subjects with Low base rate expectations had fewer hits than subjects in the control group. Base rate expectations also influenced the false alarm rate through an interaction with incentives. This interaction stems from an unexpectedly large number of false alarms in the Low base rate expectations and Detection incentive condition. On average, subjects in this condition had slightly more than seven false alarms while in the other eight experimental conditions subjects had on average fewer than one false alarm. Even so, a pattern is apparent in the results for the false alarm rate shown in Table 4. At each level of the incentives factor, subjects with Low base rate expectations had more false alarms than subjects in the control condition, and subjects with High base rate expectations had fewer false alarms than subjects in the control condition.

There is strong evidence that base rate expectations also influenced discrimination through an interaction with incentives. Both measures of discrimination are a function of the hit rate and the false alarm rate. Because of the high false alarm rate in the Low base rate expectations and Detection incentive condition, we find the poorest average discrimination for subjects in this cell of the experiment. A clear pattern emerges for  $\text{Discrimination}_{\text{BD}}$  with High base rate expectations leading to better discrimination than the control condition and Low base rate expectations leading to poorer discrimination than the control condition. A similar pattern emerges for  $\text{Discrimination}_{\text{SD}}$  although subjects with no incentive and Low base rate expectations have slightly better discrimination than subjects with no incentive who are assigned to the Control condition for base rate expectations. This difference is not statistically significant.

**Proposition 2.** *Incentives will influence performance in the detection of data errors.*

There is evidence that incentives had a direct impact on the hit rate and had an effect through the interaction with base rate expectations on the false alarm rate and on discrimination. At all levels of the base rate expectations factor, subjects with incentives to detect errors found more errors than subjects who did not have this incentive. Subjects who had incentives to conscientiously participate in the experiment (i.e., the Random incentive condition) had more hits than subjects who had no incentives. On average, subjects who had incentives to detect errors and Low base rate expectations reported significantly more false alarms and exhibited poorer discrimination than all other subjects. With the exception of these subjects, subjects with incentives to detect errors had better discrimination than subjects who did not have this incentive, and subjects who had incentives to conscientiously participate in the experiment had better discrimination than subjects who had no incentives.

## 7. Conclusion

The introduction of this paper mentions an earlier laboratory experiment in this research stream in which information-based base rate expectations had no statistically significant effect on error detection performance. Table 10 shows a comparison of the experiment reported in this paper and this earlier experiment. The two experiments used a similar experimental task, similar subjects, and the same dependent variables. The manipulation of the base rate expectations factor was different in the two experiments. In the study reported here, base rate expectations were manipulated through direct experience. In the prior experiment [5], base rate expectations were manipulated by giving subjects information about the error rate. Two levels of the incentives factors (Detection and Random) were used in both studies. The No Incentive level of the incentives factor was used for the first time in this study.

This paper provides three contributions to the literature on data quality and human performance. First, it is shown that base rate expectations affect performance in the detection of data errors when these expectations are developed

Table 10  
Comparison with an earlier experiment

Findings of this paper	Findings of an earlier experiment [5]
<p><i>Experience-based base rate expectations</i> Main effect found for hits. Effect on false alarms and discrimination through an interaction with incentives</p> <p><i>Incentives</i> Main effect found for hits. Effect on false alarms and discrimination through an interaction with incentives</p>	<p><i>Information-based base rate expectations</i> No effect found for hits, false alarms, and discrimination</p> <p><i>Incentives</i> Main effect found for hits and discrimination. No effect found for false alarms.</p>

through direct experience, although not when derived from direct information. From a theoretical standpoint, this finding provides support for the contention that a theory of error detection based on signal detection theory gives leverage for understanding the detection of data errors. As predicted by the theory, humans are shown to improve their discrimination in response to base rate expectations. The finding also contributes to the broader literature on human performance by showing that humans respond to base rates similarly in two very different types of tasks: the detection of data errors and probabilistic judgment tasks.

Second, strong confirmatory evidence for the effect of incentives on error detection performance is provided. In addition to confirming the results of an earlier experiment [5], this study provides a rich understanding of how incentives affect performance. The finding that the false alarm rate goes through the roof with strong incentives and low base rate expectations was unexpected and merits careful attention from scholars and managers. The question of whether this is a general phenomenon is not answered by this study, and the phenomenon should be confirmed in additional studies examining error detection performance as well as performance on other tasks. One might expect that an incentive system that rewards hits while penalizing false alarms would keep the false alarm rate from going through the roof. Unfortunately, prior research [5] suggests that incentive schemes that penalize false alarms may discourage error flagging activity and lead to a suppression of the hit rate.

Third, the study provides the first evidence that even in the absence of incentives, users' error detection behavior may respond to other task conditions. Even in the absence of any sort of incentive, subjects with High base rate expectations have more hits and better discrimination than subjects with Low base rate expectations.

While no single study can provide a definitive guide for practice, several guidelines are suggested by the research findings.

1. Users should be given opportunities to learn about the error rate in the data with which they work before optimal error detection performance is expected. It is possible that working closely with a mentor who has had more direct experience with the data would be useful,

although this collaboration would have to go beyond simply giving the novice user information about the error rate.

2. Managers of operations with a low error rate should proceed cautiously before giving users strong incentives to detect data errors. A careful analysis of the cost of missed errors and the cost of false alarms should precede the implementation of such an incentive program.
3. Managers of operations with a high error rate should consider implementing an incentive scheme that rewards users for detecting errors in data. The results of this study suggest that incentives to detect errors in an environment with a high error rate will improve the number of errors detected without increasing the number of false alarms.

Two directions for future research emerge from these guidelines. First, research on the question of how appropriate base rate expectations can be developed in organizational settings is needed. It is important to understand, for example, whether experience with one dataset is brought to bear on tasks done with another related dataset or whether base rate expectations are discarded when users begin working with a new dataset. Second, additional research addressing the question of how to encourage error flagging activity without allowing the number of false alarms to soar is needed. The findings of this research would be particularly useful for managers working in environments in which the cost of investigating possible data errors is high.

A limitation of this study is that the experience of working with the initial dataset with an error rate of 30% may have improved subjects' abilities to detect data errors rather than changing their expectations about the error rate. Future research should examine this possibility by training all subjects to an equivalent level of error detection competence before manipulating expectations about the base rate of errors through direct experience.

#### Appendix A. Measures of discrimination

Two measures of discrimination are used in this investigation. The first measure was developed by researchers

working in the area of signal detection theory [30,53,54]. The second was developed in the area of behavioral decision theory [55]. The measure from signal detection theory is called  $\text{Discrimination}_{SD}$  here and is discussed in Section 1. The measure from behavioral decision theory is called  $\text{Discrimination}_{BD}$  here and is discussed in Section 2.

A.1.  $\text{Discrimination}_{SD}$

Fig. 3 shows a receiver-operating characteristic (ROC) curve. The area under an ROC curve is a measure of discriminability because the extent of overlap between the noise and signal distributions shown in Fig. 4 of the body of the paper determines the area under an ROC curve [30].

All of the points on an ROC curve represent equivalent discriminability, but different response criteria. Fig. 3 shows a pair of points (a and b) with the same discriminability but with different hit and false alarm rates. The performance represented by Point a is attained by an increased willingness to report a signal compared to the performance represented by Point b [30].

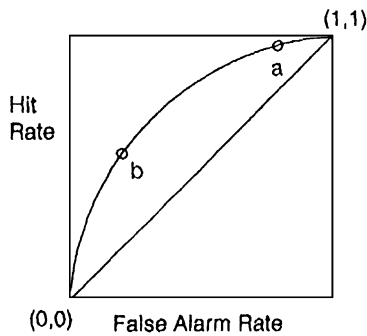


Fig. 3. Receiver-operating characteristic curve.

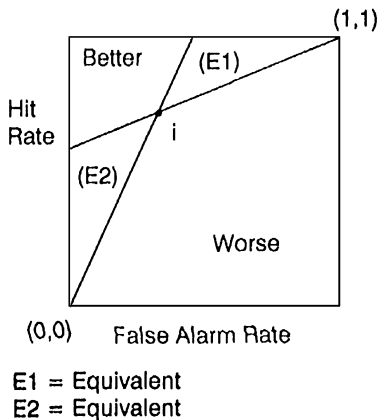


Fig. 4. Regions representing discriminability better than, worse than, and equivalent to Point i.

The area under an ROC curve can be determined if a number of points along the curve are known. This can be done by controlling discriminability across a series of experimental trials and simultaneously manipulating a subject’s response criterion. In this investigation, a within-subject manipulation of the response criterion was not made because this would have been detrimental to the critical realism of the experimental task. Instead, for each subject we have only one hit rate and one false alarm rate.

The area under an ROC curve can be estimated when we have only one point on an ROC curve. The estimate is based on determining points in a unit square representing discriminability better than the observed performance, worse than the observed performance, and equivalent to the observed performance [30,53,54].

Point i in Fig. 4 represents a known hit rate and a known false alarm rate. Fig. 4 shows the unit square divided into four regions. An ROC curve containing Point i cannot pass through the regions labeled “Better” and “Worse”. The other two regions (labeled E1 and E2) contain points that do not necessarily represent a change in discriminability compared to Point i. An ROC curve containing Point i must be drawn completely inside the regions E1 and E2 [30].<sup>1</sup>

The maximum possible area under an ROC curve that includes Point i is given by

$$\text{Maximum Area} = (E1 + E2 + \text{Worse}) / (E1 + E2 + \text{Worse} + \text{Better}).$$

The minimum possible area under an ROC curve that includes Point i is given by

$$\text{Minimum Area} = \text{Worse} / (E1 + E2 + \text{Worse} + \text{Better}).$$

$\text{Discrimination}_{SD}$  is equal to the average of the maximum possible area under the ROC curve and the minimum possible area under the ROC curve [54].

A.2.  $\text{Discrimination}_{BD}$

$\text{Discrimination}_{BD}$  is a discrimination index proposed by Yates [55]. This index is a measure of the extent to which a subject is able to correctly categorize stimuli. In this investigation, subjects worked with two types of stimuli: records containing an error and records not containing an error. Subjects placed these stimuli into two categories: records reported as errors in the confirmation memo and records not reported as errors in the confirmation memo.

If a subject were completely unable to distinguish records containing an error from records not containing an error, we would expect the number of records correctly reported as

<sup>1</sup> Additional details on the estimation procedure are given in McNicol [30].

	Item Reported as Error	Item Not Reported as Error
Error Exists	A	B
Error Does Not Exist	C	D

Fig. 5. Inputs to the calculation of  $Discrimination_{BD}$ .

containing an error in the confirmation memo divided by the total number of records reported in the confirmation memo to be equal to the base rate of errors. We would also expect the number of records containing an error but not reported in the confirmation memo divided by the total number of records not reported in the confirmation memo to be equal to the base rate of errors. The discrimination index is a measure of a subject’s improvement over this complete lack of discrimination.

Fig. 5 illustrates the definition of the four inputs (labeled A, B, C, and D) to the calculation of  $Discrimination_{BD}$ .

A is equal to the number of items reported as errors when an error exists, B is equal to the number of items not reported as errors when an error exists, C is equal to the total number of items reported as errors, and D is equal to the total number of items not reported as errors.

A three-step algorithm used to compute  $Discrimination_{BD}$  is outlined below.

*Step 1:* The number of items reported as errors that are actually errors (A) is divided by the number of items reported as errors (C). The base rate of errors in the population is then subtracted from this intermediate result. The resulting difference is first squared and then weighted by the total number of items reported as errors (C).

*Step 2:* The number of items not reported as errors that are actually errors (B) is divided by the number of items not reported as errors (D). The base rate of errors in the population is then subtracted from this intermediate result. The resulting difference is first squared and then weighted by the total number of items not reported as errors (D).

*Step 3:* The outcomes of the calculations in Step 1 and Step 2 are summed and the result is divided by the total number of items in the population (C+D).

The calculation of  $Discrimination_{BD}$  for a subject with perfect discrimination is outlined below.

*Step 1:* For a subject with perfect discrimination,  $A/C$  is equal to 1. When the base rate of errors in the population (0.20) is subtracted from this intermediate result, we get 0.80. When this difference is squared and then weighted by the total number of items reported as errors (10), we get 6.40.

*Step 2:* For a subject with perfect discrimination,  $B/D$  is equal to 0. When the base rate of errors in the population (0.20) is subtracted from this intermediate result, we get  $-0.20$ . When this difference is squared and then weighted by the total number of items not reported as errors (40), we get 1.60.

*Step 3:* When we add 6.40 and 1.60 and then divide this result by the total number of items in the population (50), we get 0.16. The upper bound of  $Discrimination_{BD}$  (i.e., the value we get for a subject with perfect discrimination) is therefore 0.16.

The calculation of  $Discrimination_{BD}$  must be modified in the extreme cases of a complete lack of discrimination. First, if no items are reported as errors (i.e., both A and C are equal to zero), Step 1 of the algorithm is skipped. Step 3 is changed to read “The outcome of the calculation in Step 2 is divided by the total number of items in the population”. Second, if all items are reported as errors (i.e., both B and D are equal to zero), Step 2 of the algorithm is skipped. Step 3 is changed to read “The outcome of the calculation in Step 1 is divided by the total number of items in the population”. In these two cases,  $Discrimination_{BD}$  is equal to zero.

**References**

- [1] Laudon KC. Data quality and due process in large interorganizational record systems. *Communications of the ACM* 1986;29(1):4–11.
- [2] Redman TC. *Data quality for the information age*. Norwood, MA: Artech House, Inc., 1996.
- [3] Paradise DB, Fuerst WL. An MIS data quality methodology based on optimal error detection. *Journal of Information System* 1991;5(1):48–66.
- [4] Orr K. Data quality and systems. *Communications of the ACM* 1998;41(2):66–71.
- [5] Klein BD, Goodhue DL, Davis GB. Can humans detect errors in data?: impact of base rates, incentives, and goals. *MIS Quarterly* 1997;21(2):169–94.
- [6] Davis GB, Neter J, Palmer RR. An experimental study of audit confirmation. *Journal of Accountancy* 1967;123(6): 36–44.
- [7] Ricketts JA. Powers-of-ten information biases. *MIS Quarterly* 1990;14(1):63–77.
- [8] Klein BD, Goodhue DL, Davis GB. Conditions for the detection of data errors in organizational settings: preliminary results from a field study. *Proceedings of the 1996 Conference on Information Quality*, 1996, p. 24–52.
- [9] Koehler JJ. The base rate fallacy reconsidered: descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences* 1996;19(1):1–53.
- [10] Davis GB, Olson MH. *Management information systems: conceptual foundations, structure, and development*. New York, NY: McGraw-Hill Book Company, 1985.
- [11] Fox C, Levitin A, Redman T. The notion of data and its quality dimensions. *Information Processing & Management* 1993;30(1):9–19.

- [12] Huh YU, Keller FR, Redman TC, Watkins AR. Data quality. *Information and Software Technology* 1990;32(8):559–65.
- [13] Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems* 1996;12(4):5–34.
- [14] Ham J, Losell D, Smieliauskas W. An empirical study of error characteristics in accounting populations. *The Accounting Review* 1985;60(3):387–406.
- [15] Johnson JR, Leitch RA, Neter J. Characteristics of errors in account receivable and inventory audits. *The Accounting Review* 1981;56(2):270–93.
- [16] Svanks MI. Integrity analysis. *Information and Software Technology* 1988;30(10):595–605.
- [17] Ballou DP, Pazer HL. Cost/quality tradeoffs for control procedures in information systems. *OMEGA: International Journal of Management Science* 1987;15(6):509–21.
- [18] Ballou DP, Pazer HL, Belardo S, Klein B. Implications of data quality for spreadsheet analysis. *Data Base* 1987;18(3):13–9.
- [19] Ballou DP, Tayi GK. Methodology for allocating resources for data quality enhancement. *Communications of the ACM* 1989;32(3):320–9.
- [20] Ballou DP, Pazer HL. Modeling data and process quality in multi-input, multi-output information systems. *Management Science* 1985;31(2):150–62.
- [21] Ballou DP, Pazer HL. Designing information systems to optimize the accuracy-timeliness tradeoff. *Information Systems Research* 1995;6(1):51–72.
- [22] Bansal A, Kauffman RJ, Weitz RR. Comparing the modeling performance of regression and neural networks as data quality varies: a business value approach. *Journal of Management Information Systems* 1993;10(1):11–32.
- [23] Gaba A, Winkler RL. Implications of errors in survey data: a Bayesian model. *Management Science* 1992;38(7):913–25.
- [24] O'Leary DE. The impact of data accuracy on system learning. *Journal of Management Information Systems* 1993;9(4):83–98.
- [25] Klein BD. How do actuaries use data containing errors?: models of error detection and error correction. *Information Resources Management Journal* 1997;10(4):27–36.
- [26] Baker EM, Schuck JR. Theoretical note: use of signal detection theory to clarify problems in evaluating performance in industry. *Organizational Behavior and Human Performance* 1975;13(3):307–17.
- [27] Ballou DP, Pazer HL. The impact of inspector fallibility on the inspection policy in serial production systems. *Management Science* 1982;28(4):387–99.
- [28] Ballou DP, Pazer HL. Process improvement versus enhanced inspection in optimized systems. *International Journal of Production Research* 1985;23(6):1233–45.
- [29] Ballou DP, Pazer HL. Designing inspection strategies for uncertain environments. *Decision Sciences* 1987;18(2):217–33.
- [30] McNicol D. A primer of signal detection theory. London: George Allen & Unwin Ltd., 1972.
- [31] Thompson SC. Detection of social cues: a signal detection theory analysis. *Personality & Social Science Psychology Bulletin* 1978;4(3):452–5.
- [32] Koppell S. Decision latencies in recognition memory: a signal detection theory analysis. *Journal of Experimental Psychology: Human Learning & Memory* 1977;3(4):445–57.
- [33] Craig A. Signal detection theory and probability matching apply to vigilance. *Human Factors* 1987;29(6):645–52.
- [34] Green DM, Swets JA. Signal detection theory and psychophysics. New York, NY: Wiley, 1966.
- [35] Pastore RE, Scheirer CJ. Signal detection theory: considerations for general application. *Psychological Bulletin* 1974;81(12):945–58.
- [36] Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240(4857):1285–94.
- [37] Davies DR, Parasuraman R. The psychology of vigilance. London: Academic Press, 1981.
- [38] Campbell JP. Modeling the performance prediction problem in industrial and organizational psychology. In: Dunnette MD, Hough LM, editors. *Handbook of industrial and organizational psychology*, 2nd ed. Palo Alto CA: Consulting Psychologists Press, 1990. p. 687–732.
- [39] Cryer EH, Bettman JR, Payne JW. The impact of accuracy and effort feedback and goals on adaptive decision behavior. *Journal of Behavioral Decision Making* 1990;3(1):1–16.
- [40] Guzzo RA, Jette RD, Katzell RA. The effects of psychologically based intervention programs on worker productivity: a meta-analysis. *Personnel Psychology* 1985;38(2):275–91.
- [41] Jenkins DC. Financial incentives. In: Locke EA, editor. *Generalizing from laboratory to field settings*. Lexington, MA: Lexington Books, 1986. p. 167–80.
- [42] Locke, EA, Feren, DB, McCaleb VM, Shaw KN, Denny AT. The relative effectiveness of four methods of motivating employee performance. In: Duncan K, Gruneberg M, Wallis D, editors. *Changes in working life*. New York, NY: Wiley, 1980. p. 363–88.
- [43] Gigerenzer G, Hell W, Blank H. Presentation and content: the use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance* 1988;14(3):513–25.
- [44] Kahneman D, Tversky A. On the psychology of prediction. *Psychological Review* 1973;80(4):237–51.
- [45] Nisbett RE, Borgida E. Attribution and the psychology of prediction. *Journal of Personality and Social Psychology* 1975;32(5):932–43.
- [46] Christensen-Szalanski JJJ, Beach LR. Experience and the base-rate fallacy. *Organization Behavior and Human Performance* 1982;29(2):270–8.
- [47] Christensen-Szalanski JJJ, Bushyhead JB. Physicians' use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance* 1981;7(4):928–35.
- [48] Manis M, Dovalina I, Avis NE, Cardoze S. Base rates can affect individual predictions. *Journal of Personality and Social Psychology* 1980;38(2):231–48.
- [49] Ginossar Z, Trope Y. Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology* 1987;52(3):464–73.
- [50] Grether DM. Bayes rule as a descriptive model: the representativeness heuristic. *The Quarterly Journal of Economics* 1980;95(3):537–57.
- [51] Kassin SM. Base rates and prediction: the role of sample size. *Personality and Social Psychology Bulletin* 1979;5(2):210–3.
- [52] Roy MC, Lerch FJ. Overcoming ineffective mental representations in base-rate problems. *Information Systems Research* 1996;7(2):233–47.
- [53] Norman DA. A comparison of data obtained with different false-alarm rates. *Psychological Review* 1965;71(1):243–6.

- [54] Pollack I, Norman DA, Galanter E. An efficient non-parametric analysis of recognition memory. *Psychonomic Science* 1964;1(1):327–8.
- [55] Yates JF. *Judgment and decision making*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1990.
- [56] Neter J, Wasserman W, Kutner MH. *Applied linear statistical models*. Homewood, IL: Irwin, 1990.
- [57] McClave JT, Benson PG. *Statistics for business and economics*. San Francisco, CA: Dellen Publishing Company, 1994.
- [58] Butt J. Frequency judgments in an auditing-related task. *Journal of Accounting Research* 1988;26:315–30.