

IMPROVING THE FORMAT OF FINANCIAL TABLES

Joseph Donelan, University of West Florida

ABSTRACT

This paper empirically examines the effects of information format on human judgment accuracy. The theoretical framework is based on the lens model of human judgment (Brunswick, 1952). An experimental design is used to test for differences in human judgment accuracy in a bond rating task. The results indicate that the use of certain table simplification techniques does not improve decision accuracy.

INTRODUCTION

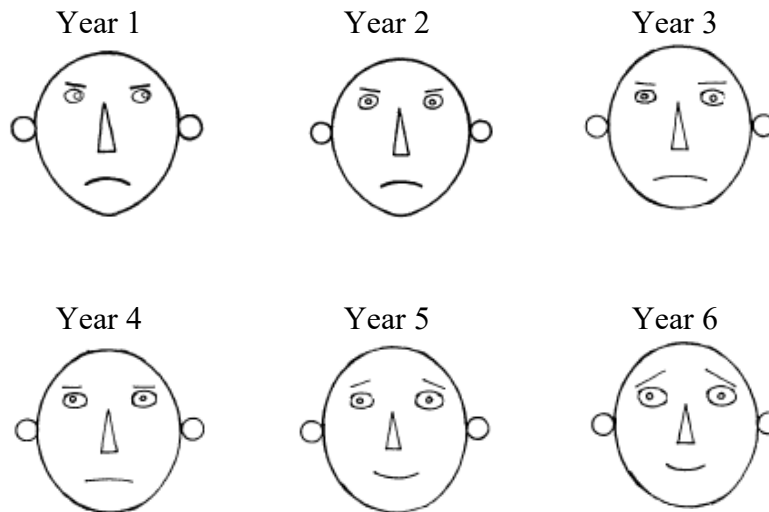
The purpose of this paper is to determine if alternative table formats result in improved human decision making. Prior studies that examined the impact of visual appearance on cognition relied on behavioral theories of human judgment and utilize laboratory designs in order to reach conclusions about the effect of changes in the data presentation on human judgment speed and accuracy. DeSanctis (1984) and Verdinelli (2013) provide a detailed review of the extant literature on the use of graphics in data presentation. For example, Benbasat and Dexter (1985) used an experimental design to study decision quality, decision-making time, and user perception when information was presented in both color-enhanced and graphic formats. Their results show that color enhanced reports were superior in some circumstances. However, there was no difference between the judgment of subjects who used graphic versus tabular formats. Vogel, Lehman and Dickson (1986) found that presentation of data using graphics as visual aids resulted in greater persuasive powers than using the same data in tabular format. But poor table construction could also be the cause of the poor persuasive power of the tabular presentation.

Stock and Watson (1984) and Chernoff (1973) tested for differences in judgment performance using two types of data presentation: multidimensional faces and tabular presentation. Subjects were asked to analyze financial data for forty-two firms using actual reported financial figures. One group was given multidimensional representations of the financial ratios of firms (see Figure 1), while a second group is given the same information in tabular form. The subjects' ability to detect changes in financial status of the firm was tested based on their ability to correctly determine if the company's bond rating was upgraded, downgraded or remained the same. The results showed that subjects using multidimensional graphics performed better than those who used tabular data. They concluded that judgment accuracy can be influenced by the format of the report.

However, because of the limitations of the laboratory design, no generalization can be made about tabular formats versus graphic formats, nor can we draw conclusions about other information formats such as simplified table formats, or multi-colored tables and charts. Therefore, additional studies are needed to provide empirical evidence which can lead us toward data presentation rules which will be useful for improving decision making.

The purpose of this paper is to determine if data presentation formats can be improved by implementing a few simple rules which make tables easier to read. The experimental design will involve a replication and extension of the Stock and Watson study.

Figure 1
Multidimensional Faces
(Source: Stock and Watson, 1984, p. 202)



CONCEPTUAL FRAMEWORK

The use of alternative presentation formats to improve decision making can be supported by the lens model of human judgment (Brunswick, 1985; Ashton, 1982). The lens model depicts the decision as a linear combination of environmental cues. In the bond rating classification problem these cues are represented by the financial ratios. Figure 2 illustrates the lens model formulation. Classification accuracy (r_a) is a function of three factors:

1. the subjects' weighting of the cues ($r_{1s} \dots r_{6s}$) in relationship to the environmental weighting ($r_{1e} \dots r_{6e}$)
2. the subjects' judgement consistency (R_s)
3. the environmental predictability (R_e)

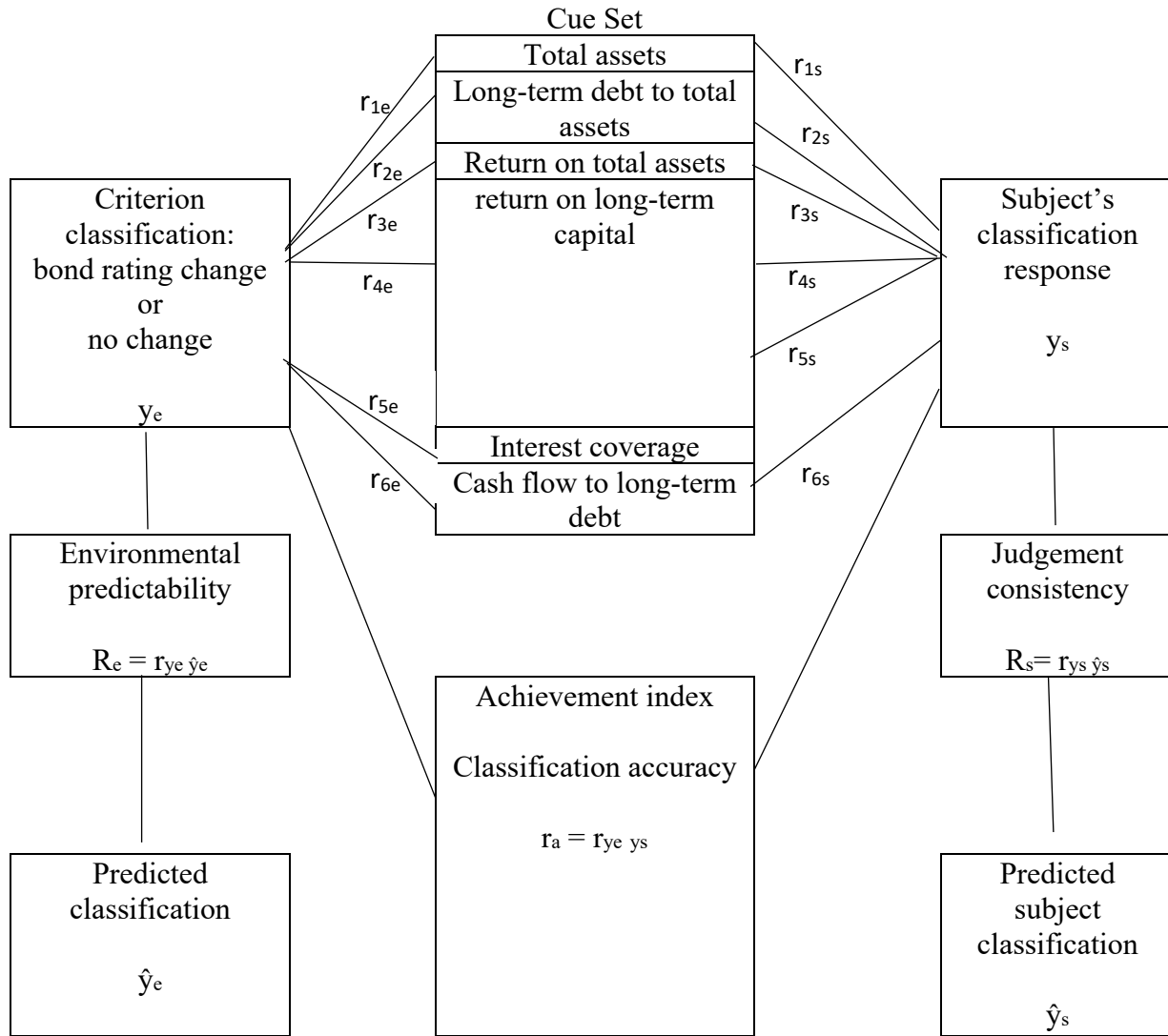
Morton (1971) suggested that graphical formats are important for data reduction and information assimilation. Therefore, improved formats represent improved cue descriptions, which should improve the subjects' ability to weight the cues and achieve consistency (R_s). Moriarity (1979) and Stock and Watson (1984) found that the use of multidimensional cue descriptions improved judgment accuracy, while other studies found opposite results (DeSanctis, 1984, p. 468).

Numerous studies have examined the effectiveness of graphical displays (see Verdinelli, 2013 for a review of the extant literature). However, very little research exists regarding numerical table design characteristics. Ehrenberg (1971) and Nicol (2013) provide basic rules for improving tabular data presentations for pattern recognition. For example,

1. round to two significant digits
2. present row and column averages

3. present important patterns in columns rather than rows
4. order rows and columns by size.

Figure 2
Lens Model Formulation of Bond Rating Decision



Although most business reports are still presented in numeric table format, the bulk of decision science research has focused on graphical displays, with a dearth of work regarding table design. To fill this void in the literature, this paper provides an empirical test of the first two rules shown above. The hypothesis is that decisions based on information presented in accordance with rules 1 and 2 (improved format) will be more accurate than decisions based on an unimproved format.

EXPERIMENTAL PROCEDURE

The experimental design is a randomized block design testing the effect of two alternative information formats on judgment quality.

The task is the identification of changes in the financial condition of firms. The financial condition is the dependent variable and is operationalized as the bond rating status. The independent variable is the treatment level which consists of two different tabular presentations of financial data for the firms. The first treatment level is the ratio data in the same tabular format used by Stock and Watson (Table 1). Data obtained from this source was compared with the example presented in Stock and Watson to determine that the ratios were computed in a similar fashion. The second treatment level is the same financial data, but in an improved tabular format (Table 2). This improved format involves rounding to two significant digits and including a column average for each column. This firms used are the same forty-two companies used in the Stock and Watson study. (A copy of the list will be made available upon request.)

Table 1						
Unimproved Format						
Year	Total Assets (Millions of Dollars)	Long-term Debt to Total Assets (%)	Return on Total Asset (%)	Return on Long-term Capital (%)	Interest Coverage (Times)	Cash Flow to Long-term Debt (%)
1	2012.232	22.564	5.770	7.000	18.010	57.005
2	2144.664	27.477	3.102	3.730	4.450	40.433
3	2153.500	25.921	4.353	5.254	5.100	48.116
4	2236.900	25.759	5.455	6.525	6.450	53.537
5	2545.300	22.760	9.362	11.551	11.570	68.839
6	2938.000	19.980	11.001	14.360	14.360	86.627

Source: Moody's Industrial Manual

Table 2						
Improved Format						
Year	Total Assets (Millions of Dollars)	Long-term Debt to Total Assets (%)	Return on Total Asset (%)	Return on Long-term Capital (%)	Interest Coverage (Times)	Cash Flow to Long-term Debt (%)
1	2000	23	6	7	18	57
2	2100	27	3	4	4	40
3	2200	26	4	5	5	48
4	2200	26	5	7	6	54
5	2500	23	9	12	12	69
6	2900	20	11	14	14	87
avg.	2300	24	7	8	10	59

Source: Moody's Industrial Manual

We hypothesize that the improved tabular format will result in improved classification accuracy.

In order to obtain a homogeneous sample, student subjects were selected from the same level of their business training – undergraduate junior and senior accounting majors currently enrolled in cost accounting. In addition, an “expert” group of nine faculty members participated.

The research instrument was administered to the students at the beginning of the class period. After the instructions were given, the participants were given thirty minutes to complete the task, and the subjects were encouraged to use all the available time.

The instructions given to the subjects were like those used by Stock and Watson. (A copy of the instructions will be made available upon request.)

RESULTS

We tested for equality of variance between the two treatment groups. This test was performed for all 42 firms combined and for the “upgrade”, “downgrade”, and “no change” categories separately. I could not reject the null hypothesis of equal variances at the .05 level in any of the comparisons. Therefore, the pooled *p* values shown in Table 3 are the appropriate statistic to test for the difference between the means of the sample.

The values shown in Table 3 represent the combined results of the student group (*n* = 41) and the "expert" group (*n* = 9). Tests were run on each group separately with no significant difference in the results.

The mean classification accuracy for the two groups is shown in Table 3. The group receiving the unimproved table format had a mean classification accuracy of 53.0% compared with the group receiving the improved format which had a mean of 51.4%. This difference is not statistically significant.

The only significant difference was in the classification accuracy of the upgraded firms. The results indicate that the group with the unimproved tables performed better in classifying these upgraded firms. But there is no apparent reason for this result. The main conclusion is that the null hypothesis of equal classification accuracy cannot be rejected therefore, it appears that the improved table format does not improve the decision accuracy.

A comparison of my results with Stock and Watson reveals several differences. First, the mean performance is 48.2% in the Stock and Watson compared with 52.1% in my study. This difference may be a result of differences in the administration of the instrument. Stock and Watson administered the instrument during the last 20 minutes of the class period and offered a single \$5 award for the best performance. We administered the instrument at the beginning of the class period and offered five awards of \$5 each, which may have resulted in higher motivation and greater concentration.

Table 3			
Mean and Standard Deviations of classification Accuracy*			
Subclass	Unimproved Tables (<i>n</i> =27)	Improved Tables (<i>n</i> =23)	Pooled <i>p</i> Value
Downgrade	61.0% (16.1)	62.9% (14.2)	.66
No Change	34.1% (16.4)	39.4% (20.6)	.31
Upgrade	65.9% (17.4)	52.8% (16.2)	.01
All Firms	53.0% (6.86)	51.4% (8.90)	.41
Classification accuracy for all participants combines 52.1%			
* Standard deviations are shown in parentheses.			

Second, our results show no significant difference in the groups based on the data presentation format. This may also be due to differences in the administration of the instrument. Students in the Stock and Watson study were motivated to perform quickly as their score was based on the number of correct classifications divided by the elapsed time. The introduction of the time constraint may explain the improved performance for the participants given the multidimensional faces. As Stock and Watson point out, cognitive theory (Glass, 1979) suggests that improved formats may facilitate mental coding, organization and recall. However, my results suggest that when time is not a constraint the improved format may lose its advantage.

There is another possible explanation for between my results and Stock and Watson's. They selected the three ratios to represent the most salient facial features of the nose length, brow angle and mouth curvature. In doing so, they removed a portion of the uncertainty involved in the decision process because the participants were not required to make a judgment themselves regarding the relative importance of the five ratios. This assistance in cue weighting may have resulted in improved performance for the group given the multidimensional faces.

One major limitation of my study is the small sample size. If there is a difference in the true population means of at least 2%, then the power ($1 - \beta$) of this experiment is 24%. That is, we have a 24% chance of finding a significant result with a sample size of 50. This limitation could easily be mitigated by doubling the sample size, which would increase the power to 42%.

Another limitation of this study is the low environmental predictability (r_e). Stock and Watson found that a discriminant model of this classification process predicts with only 47% accuracy. Although most real-life decisions also have low environmental predictability, this factor reduces the experimental control over confounding effects. Thus, random noise is increased, and the power of the test is reduced. Future research utilizing a task with greater environmental predictability would mitigate this weakness.

A third limitation is the lack of environmental validity which results from an arbitrary assignment of bond rating change probabilities. The experiment utilizes approximately equal size subgroups of upgraded, downgraded and no change firms which does not conform to the real-life probability of bond rating changes.

Also, the use of approximately equal bond rating change probabilities results in another problem in the administration of the experiment. In a similar type of experiment, Zimmer (1980) found that his results differed from prior findings by Casey (1980). In the administration of the instrument, Zimmer informed the subjects of the probability of a firm's failure/non failure classification while Casey did not reset the subjects' prior probabilities. Zimmer suggests that if the subjects' prior probabilities are not reset in the experiment, there may be systematic error in the classification process. In addition, undergraduates may have heterogeneous prior probabilities for the bond rating change task which may have reduced the power of the test.

CONCLUSION AND SUMMARY

The purpose of this paper was to provide evidence on the effects of changes in information format on human decision accuracy. The results indicate that the use of certain table simplification techniques does not improve decision accuracy. However, care should be used in interpreting these results because of the small sample size and certain internal and external validity limitations.

Research in data presentation is an important area for future research efforts. For example, future studies should examine decision tasks with greater environmental predictability. In addition,

there are many alternative tabular and graphical formats which remain untested in a classification model format.

Table 3			
Mean and Standard Deviations of classification Accuracy*			
Subclass	Unimproved Tables (n=27)	Improved Tables (n=23)	Pooled p Value
Downgrade	61.0% (16.1)	62.9% (14.2)	.66
No Change	34.1% (16.4)	39.4% (20.6)	.31
Upgrade	65.9% (17.4)	52.8% (16.2)	.01
All Firms	53.0% (6.86)	51.4% (8.90)	.41
Classification accuracy for all participants combines 52.1%			
* Standard deviations are shown in parentheses.			

REFERENCES

- Ashton, R. H. (1982). *Human Information Processing in Accounting*. Sarasota, FL; American Accounting Association
- Benbasat, I. & A. Dexter. (1985). An experimental Evaluation of Graphical Enhanced Information Presentation. *Management Science*, 31 (11), 1348-1364.
- Brunswick, E. (1952). *The Conceptual Framework of Psychology*. Chicago: University of Chicago Press.
- Casey, C. (1980). Additional Evidence on the Usefulness of Accounting Ratios for the Prediction of Corporate Failure. *Journal of Accounting Research*, 18 (2), 603-613.
- Chernoff, H. (1973). The Use of Faces to Represent Points in k-Dimensional Space Graphically. *Journal of American Statistical Association*. 15, 463-487.
- DeSanctis, G. (1984). Computer Graphics as Decision Aids: Directions for Research. *Decision Science*, 15, 463-487.
- Ehrenberg, A. (1971). Some Rules of Data Presentation. *Statistical Reporter*, 2, 305-310.
- Glass A. (1979). Holyoak, K. & J. Santa. *Cognition*. Reeding, MA: Addison-Wesley.
- Moody's Industrial Manual, www/moody's.com.
- Moriarity, A. (1979). Communication Financial Information Through Graphics. *Journal of Accounting Research*, 17 (1), 205-224.
- Morton, S. (1971). *Management Decision Systems: Computer Based Support for Decision Making*. Cambridge, MA: Harvard University.
- Nicol, A., & P. Pexman. (2010). *Displaying your findings: A practical guide for creating figures, posters, and presentations*. Washington, DC: American Psychological Association.
- Stock, D. & C. Watson. (1984). Human Judgement Accuracy, Multidimensional Graphics, and Humans versus Models. *Accounting Review*, 22 (1), 192-206.
- Verdinelli, S. & N Scagnoli. (2013). Data Display in Quantitative Research. *International Journal of Quantitative Methods*, 12, 359-381.
- Vogel, D., J. Lehman. & G. Dickson. (1986). The Impact of Graphical Displays on Persuasion: An Empirical Study. Proceedings of the Seventh International Conference in Information Systems, Society for Information Management Education Foundation, eds. L. Maggie, R. Zmund and J. Wetherbe, San Diego.
- Zimmer, I. A. (1980). Lens Study of the Prediction of Corporate Failure by Bank Officers. *Journal of Accounting Research*, 18 (3), 629-636.