

Managerial Judgment and Forecast Combination: An Experimental Study

SUNIL GUPTA

*School of Business Administration, University of Michigan, Ann Arbor, MI 48109-1234,
(313) 764-6355*

Key words: decision support, task effects, sales forecast, go/no-go decisions

Abstract

This paper examines the role of managerial judgment in forming a final forecast, or judging the achievability of a critical level of sales, when multiple forecasts or opinions are available to the decision maker. Several factors that can help improve the quality of human intervention are identified and incorporated in a decision aid. Experimental results show that aided combination can help the decision maker exploit her relevant private information and mitigate the generally observed negative effects of human intervention. Further, the results suggest that emphasizing expected sales, even when the organization is primarily interested in go/no-go decisions, helps improve performance. Several suggestions for future research are presented.

“Neither hand nor mind alone, left to themselves, amount to much;
instruments and aids are the means to perfection.” *Francis Bacon*

1. Introduction

A common managerial task facing a decision maker (DM) is to forecast the level of sales that can be expected in a future period. Often these forecasts are used as inputs for subsequent decisions, such as making a go-no-go decision by comparing the forecast sales with a predetermined critical level of sales. For example, when considering the introduction of a new product on a regional roll-out basis, forecasts for a particular region could be based on results of a test market at a representative site (not necessarily in that region), sales force surveys of the retailers in the region, and internal records about the company's sales of other products in the same region. Typically, these forecasts will not be in complete agreement. The DM must now decide what the expected level of sales may be and whether a predetermined critical level of sales can be achieved in this region.

The literature suggests that the DM who is faced with such a situation should use any one of several approaches for combining forecasts or even a simple-average heuristic. In practice, however, it is unlikely that the DM will make a final resolution based entirely on a mechanized combination rule, especially when faced with differing opinions. When considering the use of judgment in determining the combined forecast, one is confronted with the vast psychological literature on the various inconsistencies and biases in human judgments (see, for example,

Fischhoff, 1988). The overall conclusion of this growing literature is that human judgments are, on average, inferior to formal data-based analyses.

In this paper, human intervention in the forecast combination process is taken as a given. As Moriarty (1985) states, "Finding that systematic methods are superior, however, is not helpful for an organization that wishes to improve its forecast performance and yet chooses to depend on management judgment methods." Rather, the objective is to examine whether there are ways in which the advantages and disadvantages in each approach can be resolved by allowing structured interaction of judgmental and data-based combination methods. Toward this objective, this paper examines three issues that have not been sufficiently examined in the literature:

1. *Can the DM's judgment be aided to help improve (or at least not impair) the accuracy of the model-based combined forecast and the go/no-go decision?* The decision aid presented here allows the DM to modify model-based combination weights. Further, in addition to providing the outcome feedback (past outcomes and forecast errors), the decision aid also provides cognitive feedback (the modified combination weights being used by the DM and the weights derived from the data-based model), which has been found to be effective in multiple-cue probabilistic learning tasks (Balzer, Sulsky, Hammer, and Sumner, 1992).
2. *What effect does the DM's primary task – forecasting expected sales or making a go/no-go decision – have on accuracy?* Forecasts are often generated as inputs to a subsequent decision. The literature on judgment and choice suggests that the quality of human judgment can be affected by the particular task the DM's attention is focused on. Yet the existing literature on judgmental forecasting has focused exclusively on expected sales judgments.
3. *How well can the DM learn and exploit situation-specific information that has not been incorporated in the data-based model?* Although there is general pessimism about the ability of human intervention to improve model-based accuracy, some recent results in judgmental forecasting suggest that the DM can outperform statistical models when she has relevant information not incorporated in the model. The ability of the DM to properly assess the usefulness of such information and to effectively use it in *combining* forecasts has not been examined.

2. Literature review

2.1. Combination of forecasts

Various schemes for combining forecasts have been proposed (Gupta and Wilton, 1988; Moriarty, 1990). The basic formulation assumes a DM facing k forecasts, f_1 ,

f_2, \dots, f_k , of some uncertain variable y . Most approaches yield a combined forecast of the form

$$f_C = w_1f_1 + w_2f_2 + \dots + w_kf_k. \quad (1)$$

Recently, several approaches for examining the *role of human judgment* in forming combined forecasts have been suggested. For example, Blattberg and Hoch (1990) and Moriarty and Adams (1984) examine the case where the DM provides a subjective forecast that is then combined with other forecasts using a data-based approach. Results of studies using this approach have generally shown an improvement in forecast accuracy. Note that in this approach the DM, perhaps artificially, is not involved in the final combination process. Alternately, in two studies (Flores and White, 1989; Lawrence, Edmundson, and O'Connor, 1986) the DM is presented with the k constituent forecasts and forms the combined forecast judgmentally. One found an improvement; the other did not. Finally, in an untested approach recommended by Makridakis (1989) the DM is presented with the model weights, and judgment is used to *modify* these weights and arrive at the combined forecast. The decision aid described later builds on this recommendation. Here the judgmentally modified model weights are referred to as *modified weights*.

2.2. Primary task

In many forecasting situations, the primary task for the DM may not be to produce an accurate forecast of expected sales but rather to make a go/no-go decision. Yet the judgmental forecasting literature has examined almost exclusively the expected sales task. Why might this distinction in primary tasks be meaningful? The preference-reversals literature shows that there is an important distinction between judgment (preference rating or selling price of a gamble) and choice (e.g., Tversky, Sattath, and Slovic, 1988). Einhorn and Hogarth (1981) suggest that focusing on the preference judgment may lead to a more deliberative process of reasoning and evaluation of evidence. In support, Billings and Scherer (1988), in a study comparing the effects of the primary task emphasized (preference judgment or choice), found that requiring only an explicit-choice decision produced undesirable effects on information search and processing. When making the go/no-go choice, the DM *may* merely compare the various forecasts directly against the critical level of sales without going through the more deliberative process of assigning combination weights, determining the combined forecast, and *then* comparing it with the critical level of sales. The important question, in the forecast combination context, is whether focusing the DM's attention on accurate forecasts of expected sales, even when the primary interest centers on go/no-go de-

cisions, can produce more accurate go/no-go choices. The existing research in judgmental forecasting does not provide any results in this regard.

2.3. *Relevant situation-specific information*

In most laboratory studies comparing judgments with statistical models, the information available to the DM is the same as that available to the model. Consequently, as Blattberg and Hoch (1990) point out, the DMs “cannot take advantage of any skill they have at identifying other information not incorporated in the model.” In many practical forecasting tasks, on the other hand, it is possible that the DM will have access to situation-specific information not incorporated in the model, which could be relevant to improving accuracy. As Bunn and Wright (1991) caution, however, untested *conventional wisdom* can be at best irrelevant and at worst misleading. Thus, the question remains of how judgment should be incorporated in the process of combining forecasts in a manner that enables the DM to take advantage of relevant situation-specific information.

2.4. *Feedback and the decision aid*

To aid understanding of the feedback features of the decision aid, we first provide a brief description of the experimental forecasting scenario for which it was designed. Subjects were informed that they were to play the role of a marketing manager for a large manufacturer of consumer products that marketed its products in thirty districts nationwide. For introducing new products, the company could decide to do a national launch, do a roll-out launch district by district, or abort the product. To be able to make these decisions the marketing manager obtained three sales forecasts for each district based on consumer surveys and testing, sales force surveys of retailers, and internal information from company records. Given these three forecasts the marketing manager was expected to determine the levels of sales that could be expected in each of the districts and make a go/no-go decision based on whether a predetermined critical level of sales could be achieved. (The informational structure of this scenario is similar to Case 1 in Moriarty, 1990, p. 410.)

A primary function of the decision aid was to enable the DM to understand the relation between the three individual forecasts and the actual outcomes. There is general agreement in the literature that such learning can be improved by providing appropriate feedback. Two basic types of feedback have been examined in the literature: (1) *outcome feedback* provides the correct answer (the actual sales for some previously forecasted sales or whether the critical level of sales was actually achieved), and (2) *cognitive feedback* provides the DM with information that allows her to compare relations between the individual forecasts and the combined forecasts with relations in the task (Balzer et al., 1992). More specifically, model

weights and the DM's own modified weights are two instances of cognitive feedback. Much research indicates that outcome feedback is often ineffective (Balzer et al., 1992). Cognitive feedback has been found to be beneficial (Balzer et al., 1992), although there is no evidence regarding its efficacy in the forecasting or decision-making context of interest here.

The decision aid was implemented on IBM-compatible personal computers. On the main screen shown in Figure 1, the top panel showed the three component forecasts and their average. As mentioned above, equal weighting is a commonly

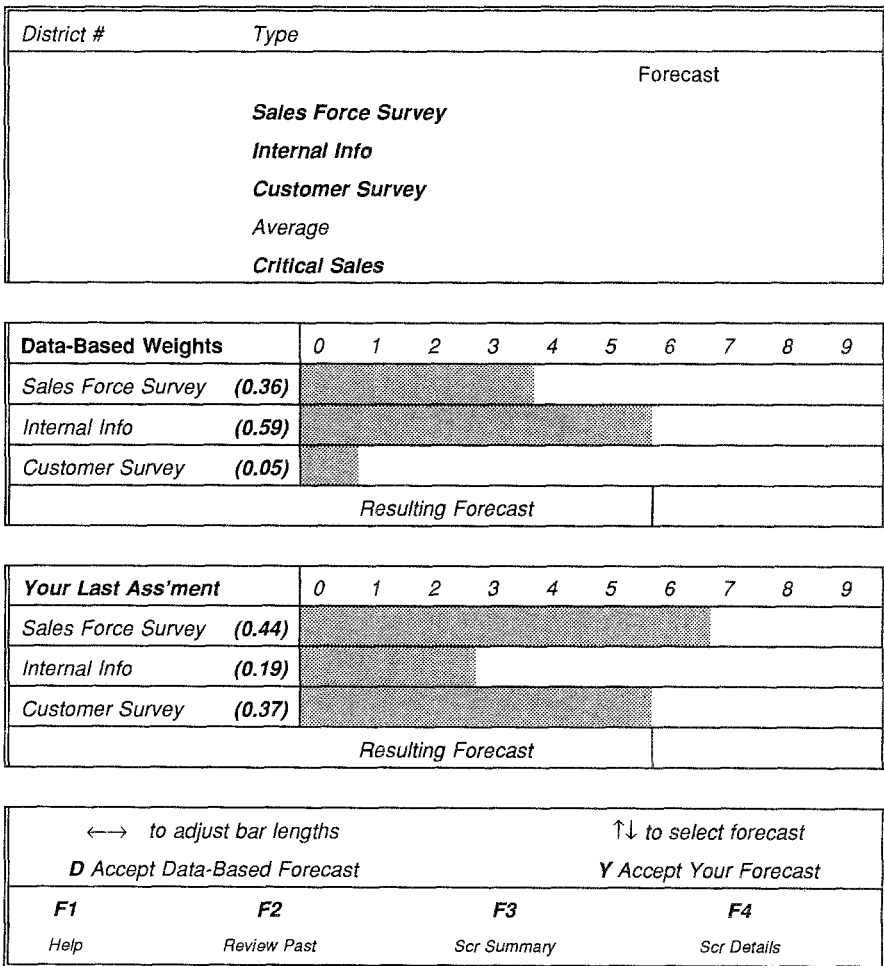


Figure 1. Main screen for aided combination

used heuristic that performs reasonably well in many forecast combination situations.

The rest of the main screen provided various types of feedback. Specifically, the middle panel of the main screen provided task-related cognitive feedback by showing the model weights, derived by using Gupta and Wilton's (1988, 1987) odds-matrix (OM) approach and the resulting combined forecast. As the subjects progressed from district to district, the weights were updated on the basis of the performance of each forecast up to that point. The OM approach was chosen because it has been shown to be particularly accurate when the available data are sparse (as is the case in the early stages of the experiment) (Gupta and Wilton, 1988, 1987) and because the derived weights are never negative, as can be the case when using an approach such as Winkler's (1981) minimum-variance method. A pretest showed that using the minimum-variance weights caused considerable confusion among the subjects. The weights were displayed using color-coded bars and numerically.

The third panel of the main screen showed three color-coded bars corresponding to the weights used by the DM for the most recent forecast, a form of subject-related cognitive feedback. Subjects could change the weights in this panel by adjusting the bar lengths. As the bar lengths were adjusted, the implied modified weights and the resulting combined forecast were continuously displayed. Finally, the DM was given the option of accepting the combined forecast based either on model weights (the middle panel) or the modified weights (the bottom panel) or entering any other value.

Outcome feedback was available through function keys (indicated at the bottom of the main screen), which allowed subjects to view a screen of histories of their own and each forecast's past errors and hits and misses in making go/no-go decisions.

3. The experiment

3.1. Manipulations

Situation specific information (info). The scenario involved three types of districts – those with low levels of expected sales (range: 2,500–9,977, average: 5,943), medium levels (range: 10,050–38,917, average: 22,967) and high levels (range: 39,523–51,657, average: 45,957). Subjects were told that actual sales had exceeded critical sales in some of the districts and fallen short in others. Therefore, the subjects had no information about whether the introduction had generally been successful or not. Finally, they were told that it was *possible*, although *not necessarily true*, that some of the input forecasts were more accurate when predicting one level of sales than another.

The relevance of this situation-specific information was manipulated by generating input forecast errors based either on a single-error covariance matrix for all

types of districts (Irrelevant) or on the basis of separate error-covariance matrices for each district type (Relevant). Because the model weights depend on the error-covariance matrix, subjects' hypothesis that the the appropriate weights changed with district type was false for the Irrelevant condition but was true for the Relevant condition. For both conditions, the model weights shown to the subjects in the middle panel of the main screen of the decision aid were not adjusted for district type. Thus, a DM could benefit from the situation-specific information by modifying weights, but only in the Relevant condition. In the Irrelevant condition, relying on this hypothesis should impair forecast accuracy. Table 1 shows the error covariance and correlation matrices, the model weights based on the OM approach, and, for comparison, Winkler's (1981) minimum-variance approach.

Feedback and decision aid. To test the efficacy of cognitive feedback and the decision aid, half the subjects used the decision aid and the other half did not. The

Table 1. Error covariance and correlation matrices and combination weights

Covariance matrix			Correlation matrix		
Irrelevant info					
45,301	42,867	12,844	1.00	0.78	0.38
42,868	66,239	7,041	0.78	1.00	0.17
12,844	7,041	25,359	0.38	0.17	1.00
Winkler weights			0.07	0.19	0.74
OM weights			0.27	0.17	0.56
Relevant info					
<i>District type: low:</i>					
286,585	232,811	11,891	1.00	0.86	0.06
232,811	256,448	56,961	0.86	1.00	0.28
11,891	56,961	160,147	0.06	0.28	1.00
Winkler weights			0.50	-0.19	0.69
OM weights			0.19	0.26	0.55
<i>District type: medium:</i>					
4,422,033	777,523	3,167,597	1.00	0.24	0.84
777,523	2,349,561	1,168,287	0.24	1.00	0.43
3,167,597	1,168,287	3,214,853	0.84	0.43	1.00
Winkler weights			0.21	0.66	0.13
OM weights			0.15	0.66	0.19
<i>District type: high:</i>					
21,054,403	266,646	5,895,217	1.00	0.01	0.35
266,646	23,165,709	15,747,536	0.01	1.00	0.89
5,895,217	15,747,536	13,440,600	0.35	0.89	1.00
Winkler weights			0.38	0.14	0.48
OM weights			0.24	0.27	0.48

procedure for those who used the decision aid to modify the model weights (Aided) has already been described. For those who did not have access to the decision aid (Unaided), the average of the three input forecasts was still provided because this is a commonly used heuristic. The outcome feedback screens were also available.

Task emphasis (task). Based on the three forecasts, the subject's task was to predict the expected level of sales *and* the go/no-go decision. However, the relative importance of these tasks was manipulated. In both cases subjects were told that they could expect to win as much as \$54, depending on the accuracy of their forecasts. The accuracy of an expected sales forecast was measured as the absolute deviation between the subject's forecast and actual sales, summed across the 30 districts. The accuracy of the achievability judgment was assessed as net hits (number of hits minus number of misses).

The desired emphasis was manipulated by changing the judgment asked for first and by altering the payoff structure. To emphasize the expected sales forecast (Sales), subjects were first asked to forecast the expected level of sales. Then the critical level of sales for the district was revealed, and they were asked to judge whether it could be achieved. Note that subjects were unaware of the critical sales when predicting the expected sales. Also, \$36 of the possible \$54 were dependent on the accuracy of the expected sales judgment. To emphasize the go/no-go decision, subjects were given the critical level of sales, asked to judge whether it could be achieved, and then asked to forecast the expected sales. In this case, a perfect score of 30 hits was worth \$36. Pre- and postexperimental checks showed that the manipulation was successful in heightening subjects' concern regarding the emphasized judgment.

3.2. *Experimental procedure*

Ten graduate students (M.B.A.s and Ph.D.s) with background in statistics or forecasting were recruited for each of the eight cells (2 Info \times 2 Decision Aid \times 2 Task) for a total of 80 subjects. Each participant was provided with printed material to study before coming for the experiment. Before starting the actual experiment, subjects were given the opportunity to go through 15 trial forecasts. The trial forecasts were created to resemble the actual error covariances as closely as possible. Subjects could continue in the trial phase as long as they needed. The trial forecasts permitted the subjects to familiarize themselves with the use of the computer, to assess whether the accuracy of the forecasts indeed changed with district types, and to form priors about the three forecasts.

At the end of the trial phase, their priors about the accuracy of the constituent forecasts in each of the three district types were elicited. Then the program proceeded to the 30 actual districts, ordered randomly. At the end of each forecast, subjects were shown their current payoff and the payoff they would receive if they

continued to forecast at the current accuracy level. Pretests showed that providing this information served to increase their involvement with the task. After forecasts and judgments had been elicited for each of the 30 districts, some postexperimental questions were asked, and then the screen displayed the final score and payoff. Subjects were paid and thanked for their participation. On average, subjects spent between one and a half to two hours to complete the task and earned \$27.43. Postexperimental tests revealed that they found the task involving.

4. Results

Tables 2 and 3 show the results for accuracy of expected sales forecasts and go/no-go predictions, respectively. The error variances of the three input forecasts in the two Info conditions – Irrelevant and Relevant – were not equal. To provide some comparability, each subject's performance is compared against what would have resulted from using the OM model weights. For the Relevant condition, the model weights used to form the baseline combined forecasts and the resulting errors were calculated on the basis of a single odds-matrix. The separate sets of weights, shown in Table 1, for high, medium, and low sales districts were not used. Thus, in the Relevant condition we simulate a situation where the DM has information that could help her outperform the baseline OM forecasts that do not utilize this information. In Table 2, the observed mean absolute error was divided by the mean absolute error that would have resulted from using the OM model

Table 2. Expected sales accuracy: Mean absolute error

Info	Task	Performance measure	Decision aid	
			Aided	Unaided
Irrelevant	Go/no-go	MAE	116.74	152.15
		Observed/OM	1.02	1.33 ^a
	Sales	MAE	116.77	133.79
		Observed/OM	1.02	1.17 ^a
Relevant	Go/no-go	MAE	1186.70	1590.96
		Observed/OM	0.86 ^b	1.15 ^a
	Sales	MAE	1134.54	1421.43
		Observed/OM	0.82 ^b	1.03

^aSignificantly worse than baseline OM accuracy.

^bSignificantly better than baseline OM accuracy.

Significant effects: Decision aid, task, decision aid*task.

Table 3. Achievability accuracy: Net hits

Info	Task	Performance measure	Decision aid	
			Aided	Unaided
Irrelevant	Go/no-go	Net hits	9.40	9.00
		Observed/OM	0.94	0.90 ^a
	Sales	Net hits	11.20	10.00
		Observed/OM	1.12 ^b	1.00
Relevant	Go/no-go	Net hits	15.40	11.20
		Observed/OM	1.10 ^b	0.80 ^a
	Sales	Net hits	16.52	11.20
		Observed/OM	1.18 ^b	0.80 ^a

^aSignificantly worse than baseline OM accuracy.

^bSignificantly better than baseline OM accuracy.

Significant effects: decision aid, task, decision aid*task.

weights. Lower MAEs and ratios smaller than one indicate greater expected sales accuracy. In Table 3, the observed net hits are compared against those resulting from using the baseline OM forecast. In this table higher net hits and ratios larger than one indicate greater go/no-go accuracy.

4.1. Does the decision-aid help improve (or at least not impair) the accuracy of the model-based combined forecast and the go/no-go decision?

Regarding forecasting accuracy, Table 2 shows that for each of the four Info and Task pairs, the Aided subjects outperformed the Unaided. Further, across the two Info conditions, only the Aided subjects were able, on average, to perform at least as well as the baseline forecast. Thus, the decision aid is helpful. However, the ability to outperform the baseline forecasts also depends on the relevance of the DM's situation-specific information. In the Irrelevant condition, the best the subjects could do was to be almost as good as model forecasts. But in the Relevant condition, with the decision aid available, they could actually outperform the less well-informed baseline forecasts.

The basic results regarding the accuracy of go/no-go decisions (Table 3) are similar to those for expected sales accuracy. For example, Aided subjects usually outperformed and, on average, did no worse than the Unaided. Further, compared to the baseline forecast, subjects with the decision aid never did significantly worse, and the improvements were most pronounced when the DM had relevant situation-specific information (the Info \times Decision Aid interaction is significant).

Together, these results imply that (1) introducing judgment into the forecast combination task can improve accuracy only when the DM has some relevant situation-specific information not available to the baseline model and (2) if judgment is going to be introduced anyway, using a decision aid such as that developed in this paper helps prevent accuracy deterioration and may even help outperform the baseline model.

4.2. What effect does the DM's primary task – forecasting expected sales or making a go/no-go decision – have on accuracy?

Table 2 shows a significant interaction between Decision Aid and Task. The task emphasis does not matter for the Aided subjects. However, for those forming Unaided combinations, emphasizing the expected Sales helps improve accuracy in both Info conditions. Table 3 shows some interesting patterns. Subjects emphasizing Sales accuracy typically were more accurate in making go/no-go decisions. Also, compared to the baseline, go/no-go decisions are most inaccurate for Unaided combination under the more complex Relevant environment. These results show that emphasizing expected sales accuracy never hurts and often helps, even when the organizational focus is primarily in go/no-go decisions. While we had speculated about this result, the existing literature had not examined this possibility.

4.3. How well can the DM learn and exploit situation-specific information that has not been incorporated in the data-based model?

The results in both tables show that the DM's ability to exploit situation-specific information depends on the availability of the decision aid. For both measures of accuracy, those subjects who had the decision aid could consistently make the adjustments needed to actually outperform the baseline model, which was not making such adjustments.

5. Discussion

The evidence from much of the past literature on judgment and decision making suggests that when models can be used, human intervention generally impairs accuracy. In this paper, by contrast, we have examined what DMs can do under more favorable conditions. Increasingly, the need for such research has been emphasized in the literature (e.g., Bunn and Wright, 1991). Our literature review identified several factors that can effect the quality of human intervention in the forecasting process. The experimental results, in agreement with the weight of past literature, showed that unaided combination can indeed be inferior to baseline model forecasts, especially when the decision maker does not have an infor-

mational advantage over the model. Perhaps more important, the results also showed that using the decision aid helped mitigate such accuracy losses and actually helped the DM exploit the private information when it was relevant. Since the DM may often be expected to be unaware of the true relevance of the private information, using a decision aid such as that developed here is highly advisable.

Much of the past literature has also ignored the importance of the particular task emphasized on the accuracy of the resulting forecasts. Our results show that emphasizing expected sales versus go/no-go decisions generally resulted in greater accuracy of both types of forecasts. Consequently, it is advisable that forecasting systems build procedures and provide incentives for the DM to provide the most accurate point forecast, even when the organizational interest centers primarily on go/no-go decisions.

Several caveats, however, apply. First, the feedback provided to subjects in this trial was immediate and of high quality. Second, the environment of the experiment was much more controlled than any real forecasting situation is likely to be. For example, the only "pressure" brought to bear on the DM was that of receiving or not receiving the explicitly stated payoffs. In most organizational contexts the relationship between performance and compensations is murkier. The informational control in the experiment was also more extreme than is likely to be the case in real situations. Our subjects had only one piece of external information to cloud (or illuminate) their judgments – whether forecast accuracy actually changed with district types. The effects of a less well-defined incentive structure and of a multitude of possible cues to adjust for are currently under study.

Our defense for these shortcomings is twofold. First, even if the added complexity reduces or eliminates the gains in comparison to the baseline methods, there seems to be little reason to expect that the advantage over unaided judgmental combinations will also be lost. Since judgment will often be used anyway, the objective of beginning the task of identifying how best to use it has been advanced. Second, given the vast amount of literature arguing against the use of human judgment, often obtained in at least as controlled an environment, the mere demonstration that accuracy can be improved through careful human intervention and appropriate task emphasis is an important result.

References

- Balzer, W.K., L.M. Sulsky, L.B. Hammer, and K.E. Sumner. (1992). "Task Information, Cognitive Information, or Functional Validity Information: Which Components of Cognitive Feedback Affect Performance?," *Organizational Behavior and Human Decision Processes* 53, 35–54.
- Billings, R.S., and L.L. Scherer. (1988). "The Effects of Response Mode and Importance in Decision Making Strategies: Judgment and Choice," *Organizational Behavior and Human Decision Processes* 41(1), 1–19.
- Blattberg, R.C., and S.J. Hoch. (1990). "Database Models and Managerial Intuition: 50% Model + 50% Manager," *Management Science* 36(8), 887–899.
- Bunn, D.W., and G. Wright. (1991). "Interaction of Judgmental and Statistical Forecasting Methods: Issues and Analysis," *Management Science* 37(5), 501–518.

- Einhorn, H.J., and R.M. Hogarth. (1981). "Behavioral Decision Theory: Processes of Judgment and Choices," *American Review of Psychology* 32, 53-88.
- Fischhoff, B. (1988). "Judgmental Aspects of Forecasting: Needs and Possible Trends," *International Journal of Forecasting* 4, 331-339.
- Flores, B.E., and E.M. White. (1989). "Subjective Versus Objective Combining of Forecasts: An Experiment," *Journal of Forecasting* 8, 331-341.
- Gupta, S., and P.C. Wilton. (1987). "Combination of Forecasts: An Extension," *Management Science* 33(3), 356-372.
- Gupta, S., and P.C. Wilton. (1988). "Combination of Economic Forecast: An Odds-Matrix Approach," *Journal of Business and Economic Statistics* 6(3), 373-379.
- Lawrence, M.J., R.H. Edmundson, and M.J. O'Connor. (1986). "The Accuracy of Combining Judgmental and Statistical Forecasts," *Management Science* 32(12), 1521-1532.
- Makridakis, S. (1989). "Why Combining Works?," *International Journal of Forecasting* 5, 601-603.
- Moriarty, M.M. (1985). "Design Features of Forecasting Systems Involving Managerial Judgments," *Journal of Marketing Research* 22 (November), 353-364.
- Moriarty, M.M. (1990). "Boundary Value Models for the Combination of Forecasts," *Journal of Marketing Research* 27 (November), 402-417.
- Moriarty, M.M., and A.J. Adams. (1984). "Management Judgment Forecasts, Composite Forecasting Models and Conditional Efficiency," *Journal of Marketing Research* 21 (August), 239-250.
- Tversky, A., S. Sattath, and P. Slovic. (1988). "Contingent Weighting in Judgment and Choice," *Psychological Review* 95(3), 371-384.
- Winkler, R.L. (1981). "Combining Probability Distributions from Dependent Information Sources," *Management Science* 27(4), 479-488.