IMPROVING JUDGMENT IN FORECASTING

Nigel Harvey
Department of Psychology, University College London

ABSTRACT

Principles designed to improve judgment in forecasting aim to minimize inconsistency and bias at different stages of the forecasting process (formulation of the forecasting problem, choice of method, application of method, comparison and combination of forecasts, assessment of uncertainty in forecasts, adjustment of forecasts, evaluation of forecasts). The seven principles discussed concern the value of checklists, the importance of establishing agreed criteria for selecting forecast methods, retention and use of forecast records to obtain feedback, use of graphical rather than tabular data displays, the advantages of fitting lines through graphical displays when making forecasts, the advisability of using multiple methods to assess uncertainty in forecasts, and the need to ensure that people assessing the chances of a plan's success are different from those who develop and implement it.

Key words: Cognitive biases, confidence, forecasting, heuristics, judgment.

The forecasting process can be divided into a number of stages (Armstrong 1985) comprising formulation of the forecasting problem, choice of method, application of method, comparison and combination of forecasts, assessment of uncertainty in forecasts, adjustment of forecasts, and evaluation of forecasts. Each of these stages may be carried out suboptimally, and each involves judgment to some extent. All of them could benefit from improved judgment.

Forecasts can be suboptimal in two ways: inconsistency and bias. People intent on improving their forecasts should minimize these components of forecast error. Inconsistency is a random or unsystematic deviation from the optimal forecast, whereas bias is a systematic one. Stewart (2001) discusses the nature of these error components in detail, but one can gain an intuitive appreciation of the difference between them from the following brief example. Given a time series of 1000 independent data points that have varied randomly around a mean value of five units, forecasts for the next 100 points should all have the value of five units. If these forecasts have an average value of five units but are scattered around that mean, they exhibit inconsistency but not bias; if they all have a value of precisely four units, they show bias but not inconsistency; if they have an average value of four units but are scattered around that mean, they contain both inconsistency and bias.

Inconsistency may arise because of variation in the way the forecasting problem is formulated, because of variation in the choice or application of a forecast method, or because the forecasting method (e.g., human judgment) itself introduces a random element into the forecast. Biases may arise automatically when certain types of judgmental or statistical methods of forecasting are applied to particular types of data series. Alternatively, they may be introduced (often unknowingly) at various stages of the forecasting process by forecasters who have stakes in particular types of outcome.

Most principles of forecasting aim to minimize inconsistency and bias at different stages of the forecasting process. Certain common strategies for achieving this are evident.

To reduce inconsistency arising from procedural variation, a number of authors argue that an effort should be made to systematize and structure various aspects of the forecasting process (e.g., Sanders and Ritzman's, 2001, principle of structuring the adjustment process; Webby, O'Connor, and Lawrence's, 2001, principle of applying structured strategies such as task decomposition when many special events have affected a series; MacGregor's, 2001, principle that some form of decomposition is generally better than none). Many principles also exploit the fact that one can reduce inconsistency by combining estimates from different sources (e.g., MacGregor's, 2001, principle of using multiple decomposition methods and multiple estimations for each one; Stewart's, 2001, principle of combining several forecasts; Armstrong's, 2001, principle that role playing should be conducted over many sessions; Wittink and Bergestuen's, 2001, principle of combining results from different methods of conjoint analysis). Authors also recognize that judgment is unreliable, and that consistency can therefore be increased by limiting its use to aspects of the forecasting process that can benefit from it (e.g., Sanders and Ritzman's, 2001, principle of mechanical integration of statistical and judgmental forecasts; Stewart's, 2001, principle of using mechanical methods to process information; Webby, O'Connor, and Lawrence's, 2001, principles of concentrating on only the most important causal forces affecting a series and of being aware that, as the amount of domain knowledge increases, one's ability to incorporate it into the forecasting process decreases).

To reduce bias, authors have developed two broad types of principle. The first is designed to lower the chances of this category of error being introduced into the forecasting process (e.g., Webby, O'Connor, and Lawrence's, 2001, principle of selecting experts who have no stake in the outcome; Sanders and Ritzman's, 2001, principle of using caution in allowing individuals to adjust forecasts when absence of bias is important; Wittink and Bergestuen's, 2001, principle of using a method of conjoint analysis (viz. magnitude estimation) that minimizes biases). The second is designed to eliminate or cancel out biases after they have been introduced (e.g., Rowe and Wright's, 2001, principle of using experts with disparate knowledge; Morwitz's, 2001, principles of adjusting intentions to avoid biases and of making allowance for the fact that errors in recalling when the last purchase was made will bias intentions for the next purchase).

Not all principles are aimed at reducing either inconsistency or bias: some are designed to tackle both these sources of error by increasing forecasters' awareness of their existence. One way of increasing such awareness is by requiring people to justify their forecasts (e.g., Stewart's, 2001, principle of requiring justification of forecasts; Rowe and Wright's, 2001, principle that Delphi feedback should include, in addition to the mean or median estimate of the panel, justification from all panelists for their separate estimates. Another way is to ensure that people receive feedback and use it to make proper evaluations of their forecasting performance (e.g. Sanders and Ritzman's, 2001, principle of documenting all judgmental adjustments and continuously measuring forecast accuracy).

DESCRIPTION OF PRINCIPLES

I have extracted from published research seven principles for improving judgment in forecasting. These principles could be incorporated into training or advice given to forecasters or into software that provides them with decision support. For each principle, I specify the stage of forecasting to which it is relevant, mention the source of error (inconsistency or bias) that it is primarily intended to reduce, and give an example of its application.

- **Use checklists of categories of information relevant to the forecasting task.**

Using checklists of relevant information relates to the problem-formulation and forecast-adjustment stages of the forecasting task. Its aim is to increase consistency in forecasts.

Forecasts for a variable may be made solely on the basis of the recent history of that variable. Often, however, the forecaster must take account of recent or expected changes in other variables. In this case, the forecaster should use a checklist of variables (or, if there are many of them, categories of variable) that past experience has shown to be relevant to the forecast.

For example, consider an editor responsible for a number of academic journals in a scientific and medical publishing firm. As part of her job, she must forecast sales (and certain other variables) for each of her journals. She could make her forecasts solely on the basis of the previous sales figures for each one. However, she would do better by taking into account a number of other factors as well. These may include agreed sales of a future special issue to a drug firm, the expected closure of a competing journal, a new campaign to increase individual subscriptions, and so on. Other types of information may appear relevant but are not. For example, a change in the editorship of a journal may have a sudden and marked effect on the number of papers submitted to it but little, if any, effect on sales.

Checklists would help because people can rarely bring to mind all the information relevant to a task when they need to do so. Their ability to search their long-term memories for such information is imperfect, and the amount they can hold in their working memories is limited. Furthermore, people are frequently influenced by information that is not relevant to their tasks. (This may be because they selectively remember occasions when some factor did influence an outcome but forget those when it did not.) Checklists can serve both to remind people of factors relevant to their forecasts and to warn them against being influenced by other categories of information.

How should checklists be compiled? The accumulated wisdom within an organization is likely to be a good starting point. In the above example, the editor's publishing manager will have more experience and be able to suggest additional factors that may affect sales. Examining past records for such contextual effects should enable the editor to determine whether the suggested factors should be included in the list. She will be looking for evidence that they produce abrupt rather than a gradual change in the sales figures.

- **Establish explicit and agreed criteria for adopting a forecast method.**

Establishing criteria for adopting a forecasting method relates to the choice-of-method and the comparison/combination-of-forecasts stages of the forecasting task. The aim is primarily to ensure procedural consistency, but it may also help to prevent individuals with stakes in particular outcomes from introducing biases.

Different forecasting methods vary in their performance with the type of data, the forecast horizon, and the error measure used. With the development of sophisticated and easy-to-use forecasting software, someone responsible for making forecasts may try out several methods for a few forecast periods and then select for future use the one that produces the best performance on some error measure. If performance of the chosen method later deteriorates, the analyst may switch to the method that is now best on that same error measure or to one that was initially best on some other error measure.

To our publishing editor and to many others who are not statistically knowledgeable but must make forecasts, this approach may have pragmatic appeal. However, there are problems with it. First, performance over a few periods is a statistically inadequate basis for selecting a method: in the presence of variability in the data, it gives no assurance that the chosen method is the best one. Second, without a costly reanalysis of the data (assuming them to be still available), there is no way of determining whether changes in the quality of forecasts are related to changes in the data, changes in the forecast method, or changes in the error measure used.

To avoid these problems, the forecaster needs to adopt explicit criteria for selecting a forecast method *before starting to forecast*. The forecaster should select an appropriate error measure (Armstrong and Collopy 1992) and decide how to choose between or combine different forecasts from the same data on the basis of the broad characteristics of the data and published research that identifies the best techniques for dealing with data having these characteristics. (Decision-support systems that do this have been incorporated into forecasting software.)

Accuracy is just one of a number of dimensions that enter into choice of forecast method. Costs of different methods must also be taken into account. Capital outlay (e.g., for software packages) is easy to assess, but training costs are more difficult to estimate (given that they are affected by the poaching of trained staff, the need for skill updating, etc.). Other factors that may be important include transparency of the forecast method to end-users, ease of providing end-users with information about the uncertainty associated with forecasts, and the speed with which the method produces the forecast. In other words, selection of a forecast method is best regarded as a multidimensional choice problem. Research suggests that people satisfice rather than optimize when making such choices (Simon 1957). Their choice is so complex that they simplify their problem. They may do this by screening out options that fail on certain critical dimensions and then accepting the first solution they find that meets the minimal criteria they have set for the other dimensions.

In an organizational context, however, the choice problem is more complex still. Different people use the same forecasts for different purposes. Some may be more willing than others to sacrifice accuracy for speed. Some may regard overforecasting as more serious than underforecasting, whereas others may hold the opposite point of view. For example, the editor responsible for forecasting her journal sales has many other tasks. She may see the forecasts as more important for how well other people perform their jobs than for how well she performs her own. In producing her forecasts, she tends to trade off accuracy for speed. However, people in the production department use the forecasts to order the paper needed to print her journals: to them, the accuracy of her forecasts is more important than the speed with which she produced them, and, furthermore, they would rather she overforecast than underforecast. In contrast, the sales people would prefer her to underforecast: if their sales fall below the forecast, they may be held to account.

Given that different individuals and departments have different criteria for an adequate forecasting method, what is the best way of proceeding? Organizational coherence is likely to suffer if analysts produce different forecasts from the same data or if managers fail to accept forecasts because they disagree with the forecasting method. From an organizational point of view, it is better for the stakeholders to make compromises before forecasting starts. In other words, all those holding a stake in the forecasting process need to agree to explicit criteria before adopting a forecasting method.

- **Keep records of forecasts and use them appropriately to obtain feedback.**

Keeping records of forecasts and using them to obtain feedback can reduce both inconsistency and bias. This principle relates to four stages of the forecasting process: choice of forecast method, application of the forecast method, combination of forecasts, and evaluation of forecasts.

People making judgmental forecasts or combining forecasts judgmentally need information that will enable them to assess their performance. This information is known as feedback. It can improve judgment and can be of various types. Outcome feedback is just information about the actual outcome for the period(s) forecast. Cognitive feedback is more highly processed information. For example, forecasters may be told they have been overforecasting by some amount over the previous 10 periods.

If forecasters are not given information about their error levels, they must derive it for themselves from a combination of outcome feedback and their original forecasts. Although forecasters usually have a record of their most recent forecast and, hence, can compare that with the corresponding outcome, they do not always keep long-term records of their previous forecasts. In the absence of this information, they must rely on their memories to estimate their overall performance over a period of time. However, memory in this situation is affected by a well-established distortion, the hindsight bias: people tend to recall their forecasts as closer to the outcome than they actually were (Fischhoff 2001).

The hindsight bias is likely to cause forecasters to overestimate the quality of their forecasts. Returning to our earlier example, the publishing editor has a record of her forecast for current sales and can compare it with the outcome. Furthermore, she has records of the previous sales figures she used, along with other information, to produce that forecast. However, she has not recorded the earlier forecasts that she made for those sales figures. Because of the hindsight bias, she will tend to recall these forecasts as closer to those figures than they actually were. As a result, she will view her overall forecasting performance more favorably than it deserves. She may then use this distorted interpretation of her performance to discount any error in her current forecast as an uncharacteristic product of her underlying forecasting ability. She would have failed to take advantage of the potential of feedback to improve this ability. She should have kept records of her forecasts and used them to obtain objective feedback about her performance.

Hindsight bias is not the only factor that may affect the evaluation of forecasts. Even when records exist, they may be used inappropriately. Evaluation that depends on searching through records may suffer from confirmation bias: people tend to search for information that confirms rather than falsifies their hypotheses. For example, in an attempt to increase sales of a new journal, the marketing department mailed promotional material to subscribers of one of the firm's existing journals. To assess the effectiveness of their campaign, the marketing personnel examined how many more subscribers of the existing journal now get the new one as well. They discovered a five percent increase over the year since the mail-shot. This is two percent more than they had forecast, and so they felt pleased with the effects of their campaign. However, because of the confirmation bias, they had failed to notice that subscribers to other comparable existing journals (that had not received the mail-shot) had also increased their subscriptions to the new journal by a similar amount. Their campaign had had no effect.

Although I have focused here on the evaluation of judgmental methods, all forecasters should keep records of their previous forecasts and use them to obtain feedback about the effectiveness of the methods they are using.

- **Study data in graphical rather than tabular form when making judgmental forecasts.**

Using graphical displays relates to the application-of-method stage of the forecasting process. It acts to reduce bias.

When people make judgmental forecasts from time series, they can study the data ir graphical form (as a set of points on a two-dimensional plot of the forecast variable agains time) or in tabular form (as a row or column of numbers). Evidence has been accumulating that forecasts from most types of series show less overall error when based on data presentec in graphical form.

Judgmental forecasts based on trended series presented graphically are much less biasec (but no more consistent) than forecasts based on the same data presented tabularly. For example, our publishing editor makes her forecasts from previous sales that are recorded as list of numbers. Sales of one journal have dropped considerably. Her forecasts for the next fev periods are likely to show a fairly consistent continuing decrease but to underestimate its rate Had the extent of her underestimation been less, she and her publishing manager might hav realized that they needed to take more drastic action than they did (e.g., cease to publish th journal rather than try to rescue it). Had the editor forecast from a graphical display of previ ous sales, she probably would have forecast sales closer to the true underlying trend in th series. She and her manager would then have been likely to act more appropriately.

- **Draw a best-fitting line through the data series when making judgmental forecast from a graphical display.**

Drawing a best-fitting line through a data series reduces inconsistency at the applicatior of-method stage of the forecasting process.

By using graphical rather than tabular displays, forecasters can reduce but not eliminate e ror in judgmental forecasts. Recent research suggests that the advantage of using graphic displays can be increased by fitting a line by eye through the data points and using this as basis for the forecasts. When data are independent and when one does not need to take caus factors into account, the line itself is a good source of forecasts. In other cases, people can b shown how to place their forecasts in relation to the line.

Thus, the publishing editor in my previous example could draw a best-fitting line throug a graph of previous sales, extend the line beyond the most recent data point, and use thi extrapolated portion of the line to obtain her forecasts.

- **Use more than one way of judging the degree of uncertainty in time-series forecasts.**

By using multiple methods, forecasters can reduce bias and inconsistency at the assessment-of-uncertainty stage of the forecasting process.

The most common way of expressing uncertainty in a forecast from a time series is to place a confidence interval around it to show the range within which there is, say, a 90 percent probability of the outcome falling. Judgmentally set intervals are typically much too narrow, indicating that people are overconfident in their forecasts (cf. Arkes 2001).

Another way to express confidence in forecasts is first to set the size of the interval and then to judge the probability that the outcome will fall within that interval. For example, the publishing editor could estimate the probability that the actual sales figure will fall within 100 units above or below her forecast. When making this type of judgment, people underestimate true probability values of greater than 50 percent and, hence, give the impression of being underconfident in their forecasts.

To get a more accurate estimate of the degree of uncertainty in forecasts, then, one could use both methods of making the judgment and average the results to reduce inconsistency and cancel out the opposing biases. Within an organization, this could be done in a number of ways, but, in general, different people should make the two judgments. The first person (say, our publishing editor) sets confidence intervals around each forecast. Intervals should correspond to high probabilities (to ensure that the second person makes probability estimates of greater than 50 percent) that vary across forecasts (so the second person does not always give the same estimate). The first person informs the second one of the size of the intervals around each forecast but not the probabilities to which they correspond—those are for the second person to estimate. The two probabilities corresponding to each interval are then averaged to produce a final estimate.

For example, the publishing editor has produced sales forecasts for three journals. She then estimates the boundaries of a 90 percent confidence interval for the first journal, a 95 percent confidence interval for the second one, and an 80 percent confidence interval for the third one. She passes her three forecasts and the three pairs of interval boundaries on to her publishing manager. The manager estimates the probabilities that the three forecasts will fall within their respective boundaries. These estimates turn out to be 70, 85, and 60 percent, respectively. The manager passes these figures back to the editor who then averages them with her original ones to produce final probability estimates (viz. 80, 90, and 70 percent) that the outcomes for the three journals will fall within the intervals that she has set.

Just averaging judgments of the same type made by different people can be expected to improve accuracy (by reducing error variance). The technique outlined above of averaging judgments of different types should produce even greater benefits by reducing bias as well. The only disadvantage is that both the interval sizes and the probabilities attributed to them will not be standardized across forecasts from different series.

- **Someone other than the person(s) responsible for developing and implementing a plan of action should estimate its probability of success.**

Different individuals should perform the planning and forecasting tasks. This reduces bias at the assessment-of-uncertainty stage of the forecasting process.

People develop and implement plans in attempts to ensure that the future will be different from what it would have been otherwise. They often need judgmental probability forecasts of a plan's success in order to decide whether to implement it and what level of resources to devote to developing contingency arrangements to put in place if it fails.

People are overconfident in their plans: they overestimate the probability that their implementation will succeed. Recently, however, it has been shown that independent assessors (e.g., consultants), while still overconfident, are not as overconfident as the originators of plans.

These findings suggest that those who develop a plan or campaign should ask someone else to make the judgmental probability forecast for its success. For example, to save a journal with a declining number of individual subscribers from closure, the publishing editor wants to go ahead with an agreement that will make it the house journal of a small learned society. If this plan succeeds, it will maintain company profits and further facilitate relations with the academic community. If it fails, the resources that would have been saved by immediate closure will be lost, and relations with the academic community (e.g., the officers of the learned society) may take a turn for the worse. Who should estimate the probability that the plan will be effective? Research suggests it should not be the publishing editor.

CONDITIONS UNDER WHICH PRINCIPLES APPLY

Forecasting depends on using information stored in human memory or in external records. The information used to make a forecast may cover just the history of the variable being forecast (univariate forecasting). Alternatively (or in addition), it may cover values in the history of one or more variables other than that for which forecasts are to be made (multivariate forecasting). The first principle (use checklists of categories of information relevant to the forecasting task) applies only to multivariate forecasting.

In applying the third principle (keep records of forecasts and use them appropriately to obtain feedback), one must bear in mind the problem of self-fulfilling prophecies (Einhorn and Hogarth 1978). In other words, a forecast may lead to an action that results in the forecast being met. An often-cited example of this is the restaurant waiter who forecasts that customers who look rich will leave larger tips than others if service is good. As a result, he provides them with better service. Not surprisingly, they then give him larger tips than other customers do. This feedback provides him with no information about the validity of his forecast.

One must take another factor into account when applying this third principle. There is a debate in the literature about the relative effectiveness of outcome feedback and cognitive feedback. There is some consensus that cognitive feedback is more useful than outcome feedback for forecasts based on many variables. The relative effectiveness of outcome feedback is greater when fewer variables are involved in producing the forecast.

Support for the fourth principle (study data in graphical rather than tabular form when making judgmental forecasts) comes primarily from research on univariate forecasting. This work suggests that the principle should be applied when data show sustained and fairly gradual trends of the sort typical of many financial and business series. When trends are extreme (e.g., exponential) or absent, the advantage of studying graphs rather than tables of data is not apparent.

The fifth principle (draw a best-fitting line through the data series when making judgmental forecasts from a graphical display) is geared to improving univariate forecasts. Research suggests that it will be particularly useful in situations of high uncertainty when data series contain high levels of random noise.

The seventh principle (someone other than the person(s) responsible for developing and implementing a plan of action should estimate its probability of success) is specific to plans of action. It does not apply to probability forecasts for the correctness of judgments about matters of fact. In other words, it concerns the effectiveness of actions rather than the correctness of views.

SUPPORT FOR THE PRINCIPLES

The research findings relevant to the principles I have proposed and the conditions under which they are assumed to apply provide stronger support for some of the principles than for others.

- **Use checklists of categories of information relevant to the forecasting task.**

Why are checklists needed? Research has shown that experts in many fields do not base their judgments on all the available relevant information and may be influenced by irrelevant factors. I shall summarize just a few of these studies.

Ebbesen and Konečni (1975) studied what information judges take into account when setting bail. They asked judges to take part in a survey, giving them eight hypothetical case records designed to simulate the information actually available in bail hearings and asking them to set bail for each one. Results showed that the judges based their judgments on the district attorney's recommendation and on the accused person's prior record and strength of local ties. Studies by the Vera Foundation (Goldfarb 1965) had shown that the most successful bail-setting strategies take strength of local ties into account. Thus, in the survey, judges indicated that their aim was to follow currently accepted best practice. However, when Ebbesen and Konečni (1975) went on to examine the information that judges actually take into account when setting bail in real court cases, they found that judges completely ignored the accused person's prior record and strength of local ties. Only the views of the district and defense attorneys and the severity of the crime influenced the level of bail set. Checklists would have prompted judges to take into account all the information that they intended to take into account.

Slovic (1969) asked stockbrokers to estimate the importance of 11 factors in their making judgments about whether to recommend stocks to clients. He also asked them to rate the strength of their recommendations for the stocks of 128 companies that varied on these factors. The influence of these factors on the recommendations did not consistently match the importance that the stockbrokers had estimated for them. Some factors they had estimated as important had virtually no influence, whereas others they had seen as only marginally important had a large effect on their recommendations.

Gaeth and Shanteau (1984) report a study of judgments of soil quality by agricultural experts. It is recognized that any material in soil other than sand, silt, and clay should be irrelevant to these judgments. Despite this, they found the experts were influenced by certain other factors (coarse fragments and moisture levels).

Evans, et al. (1995) asked doctors to estimate the importance of taking various factors into account when deciding whether to provide patients with lipid-lowering treatment. In this explicit task, the doctors demonstrated that they were aware of many of the acknowledge(risk factors for coronary artery disease (e.g., family history of the disease, evidence of arterio sclerosis, diabetes). However, when asked how likely they were to provide lipid-lowerin; treatment to various patients, these same doctors were influenced by fewer factors than the had identified as important and often by factors that they had not included in the set of im portant ones. For example, fewer than a quarter of the doctors showed evidence of taking family history of coronary artery disease and evidence of arteriosclerosis into account. Ha ries, et al. (1996) report similar results for other types of medical treatments.

Checklists have been shown to be effective in making relevant information available t those who need it. Fault trees, for example, are a type of checklist in which categories (faults are listed. They are used by those responsible for diagnosing faults in complex system The lists appear to be effective because they help people to bring to mind possibilities th; they would otherwise overlook (Dubé-Rioux and Russo 1988; Russo and Kolzow 1994).

Checklists have also been shown to be useful for improving other types of judgmen Getty, et al. (1988) developed a set of diagnostic features to help radiologists judge abno malities of the breast as either malignant or benign. Their aim was to produce a list that w; small and manageable and that included features that are largely independent of one anothe They interviewed five specialist mammographers to elicit an initial set of several dozen fe tures. They used group discussions and statistical analyses to reduce this set first to a small set of 29 features and then to a final set of 13 features. They then ran an experiment to cor pare the accuracy of diagnoses of six general radiologists with and without this checkli: Results showed that its use significantly increased the accuracy of their judgments and inde(brought their performance up to the level of the five specialist mammographers who h: participated in developing the aid.

- **Establish explicit and agreed criteria for adopting a forecast method.**

I based my arguments in favor of using explicit and agreed criteria mainly on a priori co siderations. Continual changing from one forecast method to another in an ad hoc fashi prevents the proper evaluation of any one method. Here I shall focus on evidence relevant obtaining agreement within an organization on the basis of the forecasting process.

Some parts of an organization may suffer more from overforecasting than from underfo casting. (For others, the opposite may be true.) Goodwin (1996) has pointed out that wh such asymmetric loss functions are present, forecasts are better regarded as decisions tl maximize returns for those particular sections of the organization. Problems arise when d ferent parts of the organization have different asymmetric loss functions and when the ganization as a whole has a different loss function from its component parts.

Goodwin (1996) has pointed out that use of regression techniques to debias forecasts (tained from a particular section of an organization is constrained by various factors. Fi enough historical data must be available. Second, it requires the assumption that the relati ship between outcomes, forecasts, and forecasters' use of cues is constant. (This may not reasonable: forecasters may learn to improve their performance; forecasting personnel m change; loss functions may change.) Third, forecasters who know that their estimates being corrected may put less effort into their task. Fourth, in politically sensitive envir ments, they may distort their judgments in an attempt to negate the corrections.

Given these problems and the desirability of producing forecasts that are acceptable to organization as a whole, it seems preferable for forecasters to agree on a basis of forecast before starting the process. How should they do this? In some settings, the organization m be able to reward forecasters for forecast accuracy in a manner that is not subject to asymmetric loss functions. Even then, however, the social structure of organizations may reduce the effectiveness of this strategy. For example, even though forecasters in a sales department are rewarded for their forecast accuracy, their judgments may still be influenced by departmental solidarity and by pressure from their sales-team colleagues (who are paid according to how much they exceed targets based on forecasts).

The sociotechnical approach to decision analysis has been developed to tackle situations in which individual stakeholders within an organization differ in their interpretation of its decision problems. Phillips (1982) describes a case in which a company used this approach to obtain agreement on whether to continue to manufacture an old product that might soon be banned by the government or to introduce an improved product that would beat any ban but might lose market share. Phillips (1984) describes another case in which a company used the same approach to come to an agreed decision about whether to break into a market and, if so, with what product. This approach to decision analysis may be useful for obtaining agreement about criteria for adequate forecasting. For example, it may help to encourage people to regard the forecasting process from an organizational perspective rather than from a departmental or individual perspective.

- **Keep records of forecasts and use them appropriately to obtain feedback.**

In this section, I shall first review evidence concerning the beneficial effects of outcome and cognitive feedback. Next I shall summarize studies of the hindsight bias; this distorts the recall from memory of forecasts that have not been stored as external records. Finally, I shall outline some research on the confirmation bias. This indicates that, even when forecasts are stored in external records, people have a tendency to search those records in an unbalanced way; this can result in forecasts being judged to have been more effective than they actually were.

Bolger and Wright (1994) reviewed studies of the abilities of experts in many different areas. They concluded that experts perform well when their tasks are learnable. The most crucial factor that rendered a task learnable was the immediate availability of outcome feedback.

Laboratory studies also indicate that outcome feedback is effective. Most of these experiments have employed tasks in which participants must forecast the value of a criterion variable from single values (rather than time series) of a number of predictor variables.

Schmitt, Coyle, and Saari (1977) asked people to make 50 predictions of grade point averages from hypothetical student admission scores on tests of mathematics, verbal skills, and achievement motivation. A group that received outcome feedback (information about the actual grade point averages) after each prediction performed better than a group that did not receive this information.

Fischer and Harvey (1999) asked people to combine forecasts from four different sources. These sources varied in accuracy. Hence people had to learn to weight their forecasts appropriately. A group that received outcome feedback (the actual value of a variable that had been forecast) performed significantly better than one that did not obtain feedback. The feedback group showed its advantage rapidly. However, it did not come to outperform the combined forecast obtained from a simple average of the four individual forecasts.

Outcome feedback appears to be more effective when forecasters have few predictor variables to take into account (Balzer, Doherty and O'Connor 1989). When forecasters must consider many different predictors, its effects are slow to appear or absent. This led Hammond (1971) to devise other more highly processed forms of feedback, now collectively known as cognitive feedback. They included performance feedback (e.g., information at the accuracy of judgments and biases in judgments) and details of how forecasters weigl different predictor variables relative to how they should have been weighted.

Balzer et al. (1989) reviewed the effects of various types of cognitive feedback on ju ment quality. Within the forecasting area, Murphy and Daan's (1984) study of weather fo casters is often cited as demonstrating the usefulness of this type of information. They stuc the quality of subjective probability forecasts of wind speed, visibility, and precipitation m over a year without feedback. (Forecasts were for five consecutive six-hour periods beginn zero or two hours after the forecast time.) They analyzed the data and presented results their analyses to the forecasters as feedback. They then collected a second year of data. ' weather forecasters' performance was better in the second year than it had been in the fi Murphy and Daan (1984) recognized that factors other than the provision of feedback r have contributed to this improvement; for example, the additional year of experience in pr ability forecasting may itself have facilitated performance.

Önkal and Muradoğlu (1995) studied probabilistic forecasts of stock prices and found t performance feedback led to increased accuracy and did so to a greater extent than outco feedback. Also, in the forecast combination task described above, Fischer and Harvey (19 found that providing people with updated information about the accuracy of the four indiv ual forecasters improved their judgments to a greater extent than outcome feedback alone. fact, it enabled them to outperform the combined forecast obtained from the simple aver of the four separate forecasts.

However, not all studies have found cognitive feedback to be more effective than outco feedback. Tape, Kripal, and Wigton (1992) studied probabilistic forecasting of cardiovascu death based on the presence or absence of five risk factors. Medical students first took a p test based on 40 real cases, then were trained with 173 simulated cases, and finally tool posttest based on 40 real cases taken from patient records. During the training, they receiv no feedback, outcome feedback only (viz. the correct probability of cardiac death), cognit feedback only (viz. the correct weightings of the five predictors compared with the weig ings that they had used on previous cases), or both types of feedback. In this task, traini with outcome feedback was more effective than training with cognitive feedback. T authors suggest that this pattern of results is more likely to appear in relatively straightf ward tasks where the relation between the predictors and the variable being forecast is fai simple.

This evidence suggests that it is important for forecasters to keep records of forecasts a use them to obtain feedback. Outcome feedback may be sufficient to produce improvement relatively straightforward forecasting tasks, but more highly processed feedback informati is likely to be more useful in more complex ones.

Fischhoff and Beyth (1975) were the first to identify hindsight bias. Before Preside Nixon went to Peking or Moscow in 1972, they asked students to make probabilistic for casts that the visit would have various outcomes. After the visits, they asked the students recall the probabilities they had given and to say whether they believed that each outcor had occurred. When the students believed an outcome had occurred, they were significant more likely to recall their prediction probabilities as higher than they actually were. Th hindsight bias appears to be pervasive. For example, Arkes, et al. (1981) observed it in ho pital interns and medical college faculty who made probabilistic estimates of four possib diagnoses for a case history that they read. Furthermore, the bias seems resistant to efforts eliminate it (Fischhoff 1977). Hawkins and Hastie (1990) suggested that it implies that people cannot remember previous knowledge states.

Even when records are kept, people may not use them appropriately. Work on the confirmation bias indicates that people tend to search for information that confirms rather than falsifies their hypotheses. Wason (1968) demonstrated this bias. He presented people with four cards. Each card contained a single letter or number (e.g., *A*, *B*, *2*, *3*) on the exposed side and another on the reverse side. He told them to determine the truth of the statement "All cards with a vowel on one side have an even number on the other" by indicating which (and only which) cards they would need to turn over to do so. Most people chose the card displaying *A* alone or cards *A* and *2* instead of the correct response, cards *A* and *3*. They failed to search for disconfirming evidence (i.e., card *3*).

Einhorn and Hogarth (1978) showed how Wason's (1968) findings are relevant to the sort of record-checking task under consideration here. They asked 23 statisticians to check the claim that "when a particular consultant says the market will rise . . . it always does rise" by deciding whether to observe outcomes associated with a favorable or unfavorable prediction or predictions associated with a rise or fall in the market. Specifically, they asked them to identify the *minimum* evidence needed to check the consultant's claim. Fewer than half of the responses included observing predictions associated with a fall in the market (disconfirming evidence) whereas almost all responses included observing outcomes associated with a favorable report (confirming evidence). Thus people checking records tend to look for evidence confirming their hypotheses (forecasts) but are inclined to ignore evidence that could go against them.

In summary, it is important both to keep records and to use them *appropriately* to obtain feedback about the effectiveness of forecasts.

- **Study data in graphical rather than tabular form when making judgmental forecasts.**

Angus-Leppan and Fatseas (1986) presented people with a time series as a column of 48 numbers and asked them to forecast the next 12 values. They then asked them to draw a graph of the 48 numbers and to use it to make the 12 forecasts again. Mean absolute percentage error was two percent less when data were in graphical format.

Dickson, DeSanctis and McBride (1986) asked people to make three forecasts from each of three time series. Half of the participants in the experiment saw tables of the data whereas the other half saw graphs. For eight of the nine forecasts, error levels were significantly lower when data were graphed.

Studies of judgmental forecasts of airline passenger numbers (Lawrence 1983) and economic time series (Lawrence, Edmundson and O'Connor 1985) reinforced the view that data should be presented in graphical form to maximize accuracy. Only the work of Wagenaar and Sagaria (1975) on forecasting series showing exponential growth has pointed in the opposite direction, and others have questioned the way they assessed forecasts (Jones 1979).

Harvey and Bolger (1996) investigated reasons for the advantage of graphical presentation and studied its generality. For linearly trended series, they found that error was indeed higher with tabular presentation than with graphical presentation. This was because people making forecasts underestimated the steepness of trends much more with this format than with graphical presentation. For untrended series, however, there was a slight effect in the opposite direction; error was marginally greater with graphical presentation because inconsistency (viz. scatter of forecasts around their mean or trend line) and a tendency to overforecast were somewhat higher with this format than with tabular presentation.

In summary, graphical presentation offers a clear advantage with linearly trended series, tabular presentation offers a marginal advantage with untrended series, and tabular presentation offers a disputable advantage with exponentially trended series. This suggests that graphical presentation is to be preferred as a general strategy. Only if forecasters know in advance that the series from which they will be forecasting are all, or almost all, untrended (or, perhaps, exponential) would one recommend tabular presentation.

- **Draw a best-fitting line through the data series when making judgmental forecasts from graphical displays.**

Using graphical rather than tabular displays reduces but does not eliminate error in judgmental forecasts. This error apparently arises for three reasons.

First, people use anchor-and-adjust heuristics to make forecasts. They use the last data point as a mental anchor and make some adjustment away from it to take account of whatever pattern they perceive in the series. However, in using this heuristic, people usually make insufficient adjustment (Tversky and Kahneman 1974). Two sorts of bias in judgmental forecasting have been attributed to this underadjustment: people appear to underestimate the steepness of trends and to overestimate the positivity of the first-order autocorrelation in series. A number of studies have shown these effects.

Lawrence and Makridakis (1989) asked 350 business-school students to make sales forecasts from graphs of seven-point time series of past sales. For upwardly trended series, forecasts were 4.5 percent lower than they should have been; for downwardly trended ones, they were 8.6 percent higher than they should have been. Eggleton (1982) required 100 business-administration students to make forecasts from upwardly trended and untrended series. Judgments for the trended but not the untrended series were below what they should have been, and the size of this error was greater for series with higher variance. These and many similar results (e.g., Bolger and Harvey 1993; Harvey and Bolger 1996; Sanders 1992) have been attributed to the underadjustment characteristic of people's use of anchor-and-adjust heuristics. *On average*, the last point in the data series will be on the trend line. People use this point as a mental anchor and adjust away from it to allow for the trend in the series. The observed effect occurs because their adjustments are insufficient.

Bolger and Harvey (1993) asked people to make sales forecasts from 45-point time series. They varied autocorrelation as well as trend in the series. When series were untrended, people apparently used the last data point as an anchor and adjusted away from it to take the mean level of the series into account. However, because they typically made too small an adjustment, their forecasts were too close to the last data point. As a result of such underadjustment, people give the impression that they overestimate the positivity of the first-order autocorrelation in the series.

Another source of error in judgmental forecasts is the inconsistency that people introduce into their judgments apparently to make their sequence of forecasts look like the data series. When data are independent, the sequence of forecasts should lie along the trend line in the data series. However, when Harvey (1995) asked people (who had received training in statistical regression) to make a sequence of six forecasts from graphs of 58-point time series, he found that their judgments did not lie on a trend line. Instead, they were scattered around a trend line. Furthermore, there was more random variation in the forecast sequence when there was more random variation in the data series. People making forecasts tend to be influenced by the degree of random fluctuation as well as the pattern in the data series. Of course, when someone makes a forecast for a single period, one cannot detect statistically the introduction of this randomness into the judgment. However, because error in a single forecast is as large as that in each judgment when forecasts are made for a number of periods (Harvey, Ewart and West 1997), it is reasonable to assume that it still occurs.

A final source of error in judgmental forecasts is level biases. A number of researchers have found that forecasts from untrended series are too high (e.g., Eggleton 1982; Harvey and Bolger 1996; Lawrence and Makridakis 1989) and that underestimation of downward trends exceeds that of upward ones with the same absolute slope (e.g., Harvey and Bolger 1996; Lawrence and Makridakis 1989; O'Connor, Remus, and Griggs 1997). The reason for this overforecasting is not yet clear: it may relate to people's assumptions about differences in the costs of under- and overforecasting, to expectations that external agencies are more likely to intervene if the series moves in one direction than the other (cf. Armstrong and Collopy 1993), or to wishful thinking effects.

Recent work has shown that forecasters can reduce errors from these sources by making use of a best-fitting line drawn through the data series. Alexander, O'Connor, and Edmundson (1997) have shown that a line drawn through a series of independent data points is in itself a better source of forecasts for the series than explicit forecasts made either in the presence or absence of such a line. This technique for producing forecasts implicitly would not produce good forecasts when data are not independent or when causal factors have to be taken into account. Harvey (1997) instructed people in how to use their judgment to impose a best-fitting line on a series and then how to estimate whether the data were independent or positively autocorrelated. He told them to make their forecasts on the line if they judged them to be independent and between the last point (or last forecast) and the line otherwise. This procedure reduced the error in the forecasts by half.

It is not yet clear why these techniques are effective. However, an analysis of overall error in Harvey's (1997) experiments failed to show that it was selectively reduced in trended or autocorrelated series. This suggests that the primary effect of the procedure was to decrease inconsistency rather than to reduce underadjustment. Further improvements may depend on developing better (but still simple) advice for fitting lines through data by eye.

In summary, research to date supports the recommendation that judgmental forecasters fit a line by eye through their data to use as a basis for their forecasts.

- **Use more than one way of judging the degree of uncertainty in time-series forecasts.**

Many studies have shown that people using their judgment to set, say, 95-percent confidence intervals around forecasts produce ranges that are too narrow. For example, Lawrence and Makridakis (1989) found that these intervals were about 10 percent narrower than they should have been. O'Connor and Lawrence (1989) asked people to use their judgment to set 50- and 75-percent confidence intervals around their forecasts and found that only 37.3 percent of outcomes fell within the former and just 62.3 percent of outcomes fell within the latter. O'Connor and Lawrence (1992) and Lawrence and O'Connor (1993) have obtained similar results.

This apparent overconfidence is probably another bias that arises, at least partly, from people's use of an anchor-and-adjust heuristic as a basis for their judgment (Pitz 1974; Seaver, von Winterfeldt and Edwards 1978; Spetzler and Stäel von Holstein 1975). They use the forecast as a mental anchor and set the boundaries of the interval by adjusting away from this point. However, as is usual when people use this heuristic, they make too small an adjustment (Tversky and Kahneman 1974). Hence the interval has boundaries that are too close to the forecast; its range is too narrow.

In contrast, when people estimate the probability that the actual outcome will fall within a specified range of the forecast, they underestimate probabilities that are greater than 50 percent (Harvey 1988; see also Bolger and Harvey 1995). For this type of task, people apparently use the center of the probability scale (i.e., 50%) as their mental anchor (Poulton 1989, 1994). Hence, for probabilities that are above 50 percent, the usual underadjustment from the anchor leads to judgments that are underestimates of the probabilities; people appear to be underconfident in their forecasts.

By combining both these ways of estimating uncertainty in forecasts, forecasters should able to reduce inconsistency in estimates and to cancel out biases to some extent.

- **Someone other than the person(s) responsible for developing and implementing a plan of action should estimate its probability of success.**

Harvey (1994) reviewed experiments showing that people are overconfident in their plans. A few examples must suffice here. Cohen, Dearnaley, and Hansel (1956) studied drivers' forecasts that they would be able to drive a heavy vehicle between two wooden posts. The gap between the posts was varied. For each size of gap, drivers first forecast the number of times out of five that they would be able to drive through the posts and then attempted to drive through them five times. Their forecasts exceeded their performance. Even experienced drivers estimated that they would be able to drive through a gap no wider than their vehicle on average two times out of five. Alcohol consumption increased levels of overconfidence (Cohen, Dearnaley and Hansel 1958).

Cohen and Dearnaley (1962) asked soccer players to walk towards the goal and stop when they reached a position from which they could score one, two, three, or four times out of five. They then made five attempts to score from each position. Results showed that, on average, they were about five percent overconfident about their goal-scoring performance. In other words, the average frequency of scoring from each position was about five percentage points less than they said it would be: 15 percent instead of 20 percent, 35 percent instead of 40 percent, and so on.

Overconfidence is not restricted to plans for physical actions. Harvey (1990) studied a simulated medical-treatment task. Participants had to estimate the drug dosages needed to bring a variable used for diagnosis into a range corresponding to health. After deciding on treatment, they assessed the probability of its effectiveness. Results showed that these probability forecasts were too high; people were overconfident. (The level of overconfidence was greater for more difficult versions of the task.)

More recently, Koehler and Harvey (1997, Experiment 3) and Harvey, Koehler, and Ayton (1997) used the same task to compare probability forecasts given by people who decided on the dosages with those provided by other people who had no say in determining dosages. Overconfidence was much less in those not responsible for the treatment decisions (16%) than in those who were responsible (26%). Thus, people not responsible for plans are better at estimating their likelihood of success.

IMPLICATIONS FOR PRACTITIONERS

It is important to keep records of forecasts and to use them appropriately to obtain feedback. After all, such records can be used to assess the usefulness of other principles in the chapter (and, indeed, in the book). I have been surprised at how often organizations fail to retain sufficient information about past forecasts. Management information systems should be engineered to ensure that records of previous forecasts are kept with outcome data so that people can easily compare the effectiveness of different types of forecast or forecasts from different sources. It is important to ensure that these records are well-documented and survive personnel changes and company mergers and takeovers. Organizations should regard them as part of the inheritance on which their activities depend.

Practitioners often act as informal experimenters; they try to study the effectiveness of doing things in different ways. Unfortunately, it is often difficult to make these informal investigations systematic because most organizations make many other competing demands. Undoubtedly, making such investigations more systematic would increase their effectiveness. However, organizations will provide resources to support them only if they are convinced that the benefits will outweigh the costs.

Some of the principles I (and others) propose need informal study by the organizations applying them. It is unlikely that a specific solution to a forecasting problem will work equally effectively in all organizations. Hence, in formulating some principles, I have sacrificed precision for generality. Organizations must discover for themselves how to tailor these principles to their requirements. For example, forecasters should investigate the length and composition of the checklists they use.

IMPLICATIONS FOR RESEARCHERS

Researchers have established that judgmental methods are ubiquitous in practical situations (e.g., Dalrymple 1987; Fildes and Hastings 1994; Mentzer and Cox 1984; Mentzer and Kahn 1995; Sparkes and McHugh 1984). It seems likely, however, that the increasing availability, affordability, and usability of forecasting software packages will lead to some change in this situation. The problem of combining judgment (e.g., based on knowledge of causal factors) with the output of a statistical model will then become more important. Researchers are already starting to investigate this issue. For example, Lim and O'Connor (1995) have shown that people place too much weight on their own judgmental forecasts when combining them with the output of a statistical model.

More generally, changes in forecasting requirements result in changes in the technology that supports forecasting, and these technological developments then provide a new role for judgment. In other words, technical innovations change but do not eliminate the role of judgment. Researchers respond and find out something about how well judgment performs its new role. New principles for improving judgment in forecasting are the result. There is no finite set of principles to discover; constant change in the technology supporting forecasting ensures that.

For example, currency dealers now have software support to enable them to forecast and trade on the basis of high-frequency real-time information. Traders have to use their judgment to respond quickly to profit from a situation in which many other traders have similar software. More research is needed to clarify how attentional constraints and time pressure influence this type of judgmental forecasting and decision making. This could lead to the emergence of new principles that would help both software developers and the users of current software.

CONCLUSION

Forecasts can be improved by reducing bias and inconsistency in human judgment. Principles that have been formulated for doing this generally derive from research in cognitive psychology and allied subjects but have been validated within specific forecasting contexts. However, changes in the way that practitioners operate mean that we must continually monitor the usefulness of established principles and maintain our efforts to discover new principles.

Forecasting principles are perhaps best regarded as general recommendations. In applying them in specific situations, some fine-tuning may be useful. In other words, practitioners may benefit from carrying out informal studies of their own to discover how they can best apply the principles identified by researchers within their own organizational milieus.

REFERENCES

Alexander, A., M. O'Connor & R. Edmundson (1997), "It ain't what you do, it's the way that you do it," Paper presented at International Symposium on Forecasting, Barbados.

Angus-Leppan, P. & V. Fatseas (1986), "The forecasting accuracy of trainee accountants using judgmental and statistical techniques," *Accounting and Business Research*, 16, 179–188.

Arkes, H. R. (2001), "Overconfidence in judgmental forecasting," in J. S. Armstrong, (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic.

Arkes, H. R., R. L. Wortman, P. D. Saville & A. R. Harkness (1981), "Hindsight bias among physicians weighing the likelihood of diagnoses," *Journal of Applied Psychology*, 66, 252–254.

Armstrong, J. S. (1985), *Long Range Forecasting: From Crystal Ball to Computer*. New York: Wiley. Full text at hops.wharton.upenn.edu/forecast.

Armstrong, J. S. (2001), "Role playing: A method to forecast decisions," in J. S. Armstrong, (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic.

Armstrong, J. S. & F. Collopy (1992), "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, 8, 69–80. Full text at hops.wharton.upenn.edu/forecast.

Armstrong, J. S. & F. Collopy (1993), "Causal forces: Structuring knowledge for time-series extrapolation," *Journal of Forecasting*, 12, 103–115. Full text at hops.wharton.upenn.edu/forecast.

Balzer, W. K., M. E. Doherty & R. O'Connor, Jr. (1989), "Effects of cognitive feedback on performance," *Psychological Bulletin*, 106, 410–433.

Bolger, F. & N. Harvey (1993), "Context-sensitive heuristics in statistical reasoning," *Quarterly Journal of Experimental Psychology*, 46A, 779–811.

Bolger, F. & N. Harvey (1995), "Judging the probability that the next point in an observed time series will be below, or above, a given value," *Journal of Forecasting*, 14, 597–607.

Bolger, F. & G. Wright (1994), "Assessing the quality of expert judgment," *Decision Support Systems*, 11, 1–24.

Cohen, J. & E. J. Dearnaley (1962), "Skill and judgment of footballers in attempting to score goals: A study of psychological probability," *British Journal of Psychology*, 53, 71–86.

Cohen, J., E. J. Dearnaley & C. E. M. Hansel (1956), "Risk and hazard: Influence of training on the performance of bus drivers," *Operational Research Quarterly*, 7, 67–128.

Cohen, J., E. J. Dearnaley & C. E. M. Hansel (1958), "The risk taken in driving under the influence of alcohol," *British Medical Journal*, 1, 1438–1442.

Dalrymple, D. J. (1987), "Sales forecasting practices: Results from a United States survey," *International Journal of Forecasting*, 3, 379–391.

Dickson, G. W., G. DeSanctis & D. J. McBride (1986), "Understanding the effectiveness of computer graphics for decision support: A cumulative experimental approach," *Communications of the ACM*, 20, 68–102.

Dubé-Rioux, L. & J. E. Russo (1988), "The availability bias in professional judgment," *Journal of Behavioral Decision Making*, 1, 223–237.

Ebbesen, E. & V. Konečni (1975), "Decision making and information integration in the courts: The setting of bail," *Journal of Personality and Social Psychology*, 32, 805–821.

Eggleton, I. R. C. (1982), "Intuitive time-series extrapolation," *Journal of Accounting Research*, 20, 68–102.

Einhorn, H. J. & R. M. Hogarth (1978), "Confidence in judgment: Persistence of the illusion of validity," *Psychological Review*, 85, 395–416.

Evans, J. St. B. T., C. Harries, I. Dennis & J. Dean (1995), "General practitioners' tacit and stated policies in the prescription of lipid-lowering agents," *British Journal of General Practice*, 45, 15–18.

Fildes, R. & R. Hastings (1994), "The organization and improvement of market forecasting," *Journal of the Operational Research Society*, 45, 1–16.

Fischer, I. & N. Harvey (1999), "Combining forecasts: What information do judges need to outperform the simple average?" *International Journal of Forecasting*, 15, 227–246.

Fischhoff, B. (1977), "Perceived informativeness of facts," *Journal of Experimental Psychology: Human Perception and Performance*, 3, 349–358.

Fischhoff, B. (2001), "Learning from experience: Coping with hindsight bias and ambiguity," in J. S. Armstrong, (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic.

Fischhoff, B. & R. Beyth (1975), "'I knew it would happen.' Remembered probabilities of once-future things," *Organizational Behavior and Human Decision Processes*, 13, 1–16.

Gaeth, G. J. & J. Shanteau (1984), "Reducing the influence of irrelevant information on experienced decision makers," *Organizational Behavior and Human Performance*, 33, 263–282.

Getty, D. J., R. M. Pickett, S. J. D'Orsi & J. A. Swets (1988), "Enhanced interpretation of diagnostic images," *Investigative Radiology*, 23, 244–252.

Goldfarb, R. L. (1965), *Ransom.* New York: Harper & Row.

Goodwin, P. (1996), "Subjective correction of judgmental point forecasts and decisions," *Omega*, 24, 551–559.

Hammond, K. R. (1971), "Computer graphics as an aid to learning," *Science*, 172, 903–908.

Harries, C., J. St. B. T. Evans, I. Dennis & J. Dean (1996), "A clinical judgment analysis of prescribing decisions in general practice," *Le Travail Humain*, 59, 87–111.

Harvey, N. (1988), "Judgmental forecasting of univariate time-series," *Journal of Behavioral Decision Making*, 1, 95–110.

Harvey, N. (1990), "Effects of difficulty on judgmental probability forecasting of control response efficacy," *Journal of Forecasting*, 9, 373–387.

Harvey, N. (1994), "Relations between confidence and skilled performance," in G. Wright & P. Ayton, (eds.), *Subjective Probability.* New York: Wiley, pp. 321–352.

Harvey, N. (1995), "Why are judgments less consistent in less predictable task situations?' *Organizational Behavior and Human Decision Processes*, 63, 247–263.

Harvey, N. (1997), "Improving judgmental forecasts," Paper presented at International Symposium on Forecasting, Barbados.

Harvey, N. & F. Bolger (1996), "Graphs versus tables: Effects of presentation format on judgmental forecasting," *International Journal of Forecasting*, 12, 119–137.

Harvey, N., T. Ewart & R. West (1997), "Effects of data noise on statistical judgment," *Thinking and Reasoning*, 3, 111–132.

Harvey, N., D. J. Koehler & P. Ayton (1997), "Judgments of decision effectiveness: Actor-observer differences in overconfidence," *Organizational Behavior and Human Decision Processes*, 70, 267–282.

Hawkins, S. A. & R. Hastie (1990), "Hindsight: Biased judgments of past events after outcomes are known," *Psychological Bulletin*, 107, 311–327.

Jones, G. V. (1979), "A generalized polynomial model for perception of exponential growth," *Perception and Psychophysics*, 25, 232–234.

Koehler, D. J. & N. Harvey (1997), "Confidence judgments by actors and observers," *Journal of Behavioral Decision Making*, 10, 117–133.

Lawrence, M. J. (1983), "An exploration of some practical issues in the use of quantitative forecasting models," *Journal of Forecasting*, 2, 169–179.

Lawrence, M. J., R. H. Edmundson & M. J. O'Connor (1985), "An examination of the accuracy of judgemental extrapolation of time series," *International Journal of Forecasting*, 1, 25–35.

Lawrence, M. J. & S. Makridakis (1989), "Factors affecting judgmental forecasts and confidence intervals," *Organizational Behavior and Human Decision Processes*, 42, 172–187.

Lawrence, M. & M. O'Connor (1993), "Scale, variability and the calibration of judgmental prediction intervals," *Organizational Behavior and Human Decision Processes*, 56, 441–458.

Lim, J. S. & M. O'Connor (1995), "Judgmental adjustment of initial forecasts: Its effectiveness and biases," *Journal of Behavioral Decision Making*, 8, 149–168.

MacGregor, D. G. (2001), "Decomposition for judgmental forecasting and estimation," in J. S. Armstrong, (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic.

Mentzer, J. T. & J. E. Cox, Jr. (1984), "Familiarity, application and performance of sales forecasting techniques," *Journal of Forecasting*, 3, 27–36.

Mentzer, J. T. & K. B. Kahn (1995), "Forecasting technique familiarity, satisfaction, usage and application," *Journal of Forecasting*, 14, 465–476.

Morwitz, V. G. (2001), "Methods for forecasting from intentions data," in J. S. Armstrong, (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic.

Murphy, A. H. & H. Daan (1984), "Impacts of feedback and experience on the quality of subjective probability forecasts: Comparison of results from the first and second years of the Zierikzee experiment," *Monthly Weather Review*, 112, 413–423.

O'Connor, M. & M. Lawrence (1989), "An examination of the accuracy of judgmental confidence intervals in time series forecasting," *Journal of Forecasting*, 8, 141–155.

O'Connor, M. & M. Lawrence (1992), "Time series characteristics and the widths of judgmental confidence intervals," *International Journal of Forecasting*, 7, 413–420.

O'Connor, M., W. Remus & K. Griggs (1997), "Going up-going down: How good are people at forecasting trends and changes in trends?" *Journal of Forecasting*, 16, 165–176.

Önkal, D. & G. Muradoğlu (1995), "Effects of feedback on probabilistic forecasts of stock prices," *International Journal of Forecasting*, 11, 307–319.

Phillips, L. D. (1982), "Requisite decision modelling: A case study," *Journal of the Operational Research Society*, 33, 303–311.

Phillips, L. D. (1984), "A theory of requisite decision models," *Acta Psychologica*, 56, 29–48.

Pitz, G. F. (1974), "Subjective probability distributions for imperfectly known quantities," in L. W. Gregg, (ed.), *Knowledge and Cognition.* New York: Wiley.

Poulton, E. C. (1989), *Bias in Quantifying Judgments.* Hove: Erlbaum.

Poulton, E. C. (1994), *Behavioural Decision Theory: A New Approach.* Cambridge: Cambridge University Press.

Rowe, G. & G. Wright (2001), "Expert opinions in forecasting: The role of the Delphi technique," in J. S. Armstrong, (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic.

Russo, J. E. & K. Y. Kolzow (1994), "Where is the fault in fault trees?" *Journal of Experimental Psychology: Human Perception and Performance*, 20, 17–32.

Sanders, N. R. (1992), "Accuracy of judgmental forecasts: A comparison," *Omega*, 20, 353–364.

Sanders, N. R. & L. P. Ritzman (2001), "Judgmental adjustment of statistical forecasts," in J. S. Armstrong, (ed.), *Principles of Forecasting.* Norwell, MA: Kluwer Academic.

Schmitt, N., B. W. Coyle & B. B. Saari (1977), "Types of task information feedback in multiple-cue probability learning," *Organizational Behavior and Human Performance*, 18, 316–328.

Seaver, D. A., D. von Winterfeldt & W. Edwards (1978), "Eliciting subjective probability distributions on continuous variables," *Organizational Behavior and Human Performance*, 21, 379–391.

Simon, H. (1957), *Administrative Behavior.* New York: Wiley.

Slovic, P. (1969), "Analyzing the expert judge: A descriptive study of a stockbroker's decision processes," *Journal of Applied Psychology*, 53, 255–263.

Sparkes, J. R. & A. K. McHugh (1984), "Awareness and use of forecasting techniques in British industry," *Journal of Forecasting*, 3, 37–42.

Spetzler, C. S. & C. A. S. Stäel von Holstein (1975), "Probability encoding in decision analysis," *Management Science*, 22, 340–358.
Stewart, T. R. (2001), "Improving reliability of judgmental forecasts," in J. S. Armstron (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic.
Tape, T. G., J. Kripal, & R. S. Wigton (1992), "Comparing methods of learning clinical prediction from case studies," *Medical Decision Making*, 12, 213–221.
Tversky, A. & D. Kahneman (1974), "Judgment under uncertainty: Heuristics and biase *Science*, 185, 1127–1131.
Wagenaar, W. A. & S. D. Sagaria (1975), "Misperception of exponential growth," *Perception and Psychophysics*, 18, 416–422.
Wason, P. C. (1968), "Reasoning about a rule," *Quarterly Journal of Experimental Psychology*, 20, 273–281.
Webby, R., M. O'Connor & M. Lawrence (2001), "Judgmental time-series forecasting using domain knowledge," in J. S. Armstrong, (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic.
Wittink, D. R. & T. Bergestuen (2001), "Forecasting with conjoint analysis," in J. S. Ar strong, (ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic.