The Journal of Psychology and Financial Markets
2002, Vol. 3, No. 3, 152–160



# Expert Judgments: Financial Analysts Versus Weather Forecasters

**Tadeusz Tyszka and Piotr Zielonka**

*Two groups of experts, financial analysts and weather forecasters, were asked to predict corresponding events (the value of the Warsaw Stock Exchange Index and the average temperature of the next month). When accounting for inaccurate judgments, we find that weather forecasters attach more importance to probability than financial analysts. Although both groups revealed the overconfidence effect, it was significantly higher among financial analysts. These results are discussed from the perspective of learning from experience.*

## Expert Judgments

This paper examines what importance experts attach to various justifications of their forecasting failures. In particular, we focus on justifications appealing to the probabilistic nature of forecast events. We try to determine the consequences for experts' self-confidence when they are aware of the probabilistic nature of event forecasting. Finally, we want to learn how all these things are influenced by the type of method used to make expert judgments.

For obvious reasons, research on expert judgments focuses on accuracy. Although accuracy is what justifies asking experts for their opinions, much research shows that, in some domains, experts are not doing very well. For example, research by Tetlock [1999] on predictions of political events finds that political experts are only *slightly* more accurate than chance.

A direct evaluation of the accuracy of expert judgments may be difficult or impossible, so researchers often study the stability of expert judgments and the consensus between experts. This kind of research indicates considerable differences between experts from different domains. For example, Shanteau [1995] quotes findings on internal consistency for medical pathologists from 0.40 to 0.50, and for auditors from 0.83 to 0.90. Similarly, interjudge correlations for stockbrokers and clinical psychologists were below 0.40, and for auditors around 0.70. This illustrates that experts in some domains are doing relatively better than in others.

**Tadeusz Tyszka** is the head of the Centre for Market Psychology of Leon Kozminski Academy of Entrepreneurship and Management, and a Professor of Psychology at the Institute of Psychology of Polish Academy of Sciences. His main field of research is judgement and decision making, and economic psychology. He has published numerous articles and books in the field.

**Piotr Zielonka** is the member of the Centre for Market Psychology of Leon Kozminski Academy of Entrepreneurship and Management. He does research in behavioral finance and philosophy of science.

Requests for reprints should be sent to: Tadeusz Tyska, Centre for Market Psychologyof Leon Kozminski Academy of Entrepreneurship and Management, Warszawa, Poland. Email: tt@psychpan.waw.pl

Studies on expert judgments also show that, similarly to laypeople, experts are subject to many biases, particularly what is referred to as the overconfidence effect. In so-called probability calibration research, the probability assessments of a judge are compared with the proportion of true responses. As Lichtenstein, Fischhoff, and Philips [1982] show, it is typically found that the proportion of true responses is much smaller than the probability assigned to various propositions. This means that people generally think they know more than they actually do. And this is also the case with experts.

For example, as Tetlock [1999] showed, experts assigned extremely high confidence estimates to their predictions, although they rarely exceeded chance predictive accuracy. Moreover, the experts were much more overconfident than what is customarily observed in confidence calibration research. Experts who assigned confidence estimates of 80% or higher were correct in only 45% of cases. As noted by Korn and Laird [1999], a high level of overconfidence seems to be characteristic for financial analysts, and probably for many other kinds of experts as well.

But how do experts react to the confirmation or disconfirmation of their forecasts? This is a crucial question, because their reactions may determine whether they are capable of learning from experience. And it is well-known that the very definition of an *expert* includes the idea of using one's experience. According to Webster's (1979) dictionary, *expert* means *having, involving, or displaying special skill or knowledge derived from training or experience*.

Do experts really learn from experience? In many cases, they unquestionably do. But when it comes to forecasting and learning about probabilistic relationships between events, the answers are less clear. One question concerns cognitive ability to learn from probabilistic relationships. Brehmer [1980], who studied people's abilities to learn from experience in probabilistic situations, concluded that people tend to perceive relationships between variables as deterministic rather than

probabilistic, and that they have a number of biases that prevent them from learning from experience.

The forecasts formulated by experts also involve other kinds of factors that can hinder learning from experience. The information about the failure of an expert's forecast concerns something important to the individual – his/her reputation or self-esteem. As research on self-assessment shows, information about performance activates an important motivation to maintain a positive self-evaluation (Tesser and Campbell, 1983). People simply *want* to believe they are competent; they want others to believe they are competent, and they will behave in such a way as to maintain positive feelings about their abilities (see Tesser and Campbell, 1983, p. 5). But under some conditions, and even under risk of damage to their self-esteem, there can be another motivation: obtaining an *accurate* self-assessment (Trope, 1983).

Corresponding to these two kinds of motivation, we can expect to see two different attitudes in experts when they receive information about the failure of their forecasts. One attitude is to defend one's own self-esteem. Tetlock [1999], in his research on political experts, illustrated several strategies used by experts to defend their self-esteem. One of the most popular was the close-call counterfactual claim that the predicted outcome "almost occurred." Another argument was that some unexpected event had occurred that changed the "fundamental forces" on which the forecasts were initially predicated.

An alternative way of accounting for inaccurate forecasts can be attentiveness to factors influencing the events. For example, a weather forecaster whose forecast failed can start to look for the causes of the errors – e.g., those related to the probabilistic nature of the events predicted.

In attempting to determine when each of these attitudes will dominate, the following questions arise.

- How do experts account for failures of their forecasts?
- What is the impact of the success or failure on experts' subsequent self-evaluations – do experts whose predictions were confirmed versus those whose predictions failed sustain varied levels of self-evaluation?
- What is the impact of the success or failure on experts' subsequent confidence judgments – do experts increase their confidence in their subsequent predictions after success and decrease it after a failure?

### Clinical versus Actuarial Judgments

Dawes, Faust, and Meehl [1989] introduced a very important distinction between two methods of making expert judgments (including predictions). One method, which they call clinical (after its common use in clinical settings), collects the information the expert possesses in his or her head. The contrasting method is called actuarial (after its use in life insurance agencies) or statistical. In this method, judgments or predictions are made based on an external procedure that reflects the empirically established relationships between the events of interest. Indeed, experts predicting events under conditions of uncertainty in some domains (meteorology, econometrics) use special statistical formulas (algorithms), while experts in other domains (stockbrokers, clinical psychologists, political scientists) use what Dawes, Faust, and Meehl [1989] refer to as clinical judgments.

Weather forecasting and financial forecasting are good examples of the two methods. A variety of forecasting techniques are available to weather forecasters. The climatology method, involving weather statistics accumulated over many years, is the most elemental. No theory is needed, because the predictions are based exclusively on data analysis. Climatology, however, only works properly when the weather pattern is similar to that expected for the chosen time of year.

Another is Numerical Weather Prediction (NWP), which uses theoretical forecast models run on computers to predict a particular atmospheric factor (such as temperature) based on dozens of input variables. Common to these two methods is that the result is obtained from a kind of mathematical formula.

By contrast, in forecasting stock prices, a financial analyst implements one or a combination of two methods of market analysis —fundamental or technical analysis—to predict stock prices in the future.

It is natural to believe that experts in domains where statistical procedures are used have a better chance of correctly assessing the statistical nature of the events predicted than those from more clinical domains. This should be especially true in such domains as weather forecasting, because there are several statistical models of forecasts available. Such experts may be aware that success is more probable over repeated predictions, and that a single prediction always poses a greater chance of being wrong.

Presumably, this type of experience is less likely in domains where clinical judgments are used, where there are few systematic observations of empirical relationships. But what are the consequences of such varied experiences?

First, in explaining their inaccurate forecasts, we assume that experts in more statistical domains would tend to blame probability more often than experts from clinical domains. Correspondingly, we formulated Hypothesis 1:

> When accounting for inaccurate judgments, experts in domains where statistical procedures are used should attach more importance to the proba-

bilistic nature of the events predicted and, therefore, to probabilistic arguments, than experts from domains where clinical judgments are used.

Second, we assume that the process of thinking about why a forecast might fail—including the probabilistic nature of the events predicted—should result in lower self-evaluations in experts from statistical domains than in experts from clinical domains. Correspondingly, we formulated Hypothesis 2:

> Motivating experts to think about the reasons why a forecast might fail should result in a lower level of self-evaluation in experts in domains where statistical procedures are used, but not in experts from domains where clinical judgments are used.

Finally, we expect the well-established overconfidence effect to occur in all groups of experts. However, experts in statistical domains should be generally less overconfident, because they would tend to be more aware of the probabilistic nature of events than experts from clinical domains. Correspondingly, we formulated Hypothesis 3:

> When forecasting uncertain events, experts in domains where statistical procedures are used should manifest less overconfidence than experts from domains where clinical judgments are used.

## Method

### Respondents

Within the two groups of experts, we assume that financial analyst forecasts are based mainly on clinical judgments, while the weather forecasters use more statistical models. (Actually, one of the weather forecasters told us of a saying that high-quality weather forecasts should be made with the curtains drawn.)

### Predicted Events

Financial analysts were asked to forecast the value of the Warsaw Stock Exchange Index (WIG) at the end of 2000 (about one and one-half months from the date the research was run). Weather forecasters were asked to predict the average temperature in April 2001 (the research was run in mid-March).

In both cases, three mutually exclusive and exhaustive alternatives (events) were specified in such a way that each event approximated a 0.333 chance of occurrence. For financial analysts there were three intervals of the WIG index: below 15,500, from 15,500 to 17,000, and above 17,000. These were arbitrarily created to approximate equal-probability intervals. For weather forecasters, we specified three intervals of temperature–below 7 degrees centigrade, from 7 to 9 degrees centigrade, and above 9 degrees centigrade–each of which included 33.3% of the average temperature in April in the last fifty years.[1]

### Procedure

Two sessions were run. During Session 1, respondents were asked to assess on a 9-point Likert scale how confident they were in their assessment of the underlying forces that shape the WIG index or the average temperature in April. Then they were asked to rank-order their respective three events (concerning the WIG index or the average temperature) from the most to the least likely (ties were allowed). Finally, respondents assigned a subjective probability (level of confidence) to each of the three events. (Experts were asked to make sure that the subjective probabilities they assigned to each scenario summed to 100%.)

After about two months (when the specified forecasting interval had elapsed), we contacted the respondents again for Session 2. Respondents were first told whether their forecasts had been right or wrong. Regardless of the outcome, each respondent was presented with a list of reasons why a forecast might fail. These lists were not identical for both groups, but shared some items. (The lists were prepared in advance on the basis of focus group interviews with financial analysts and weather forecasters.) Respondents were asked to mark on a 7-point scale the importance of each reason (from 1 for "not important at all," to 7 for "very important").

Then respondents were given another questionnaire in which they were asked the same questions as in Session 1, i.e., self-assessments as experts, and to assign a subjective probability to each of three new events. We had the financial analysts predict the value of the WIG index in mid-June 2001 and the weather forecasters predict the average temperature in June 2001. The main aim of the questionnaires was not to estimate the forecast abilities of the respondents but to examine their self-evaluation, probability assessments, and justifications of forecast failure.

## Results

One-third of the financial analysts and approximately two-thirds of the weather forecasters were correct in their forecasts. We attribute the difference to the fact that weather forecasters frequently marked two or three intervals as equally probable. If one of two marked intervals was correct, the respondent was considered successful. The same procedure was followed for those who marked all three intervals as equally probable.

Note that none of the results we discuss next are dependent on the accuracy of the prediction.

Figures 1 and 2 show the results of Session 1: 1) how experts evaluated their knowledge of the underlying forces that shape the event in question (self-evaluation), and 2) the subjective probability assigned to the most probable alternative. As Figure1 shows, mean self-evaluations in the two groups of experts were almost identical, equaling 6 on a 9-point scale. Thus, both groups maintained a rather positive self-evaluation.

Figure 2 indicates that both groups revealed the overconfidence effect. As mentioned above, the alternative events were specified in such a way that each approximated a 0.33 chance of occurrence. But the confidence of experts in both groups was much higher than that. However, in agreement with Hypothesis 3, the

**FIGURE 1**
**Mean Self-Evaluations of Two Groups of Experts for Session 1**

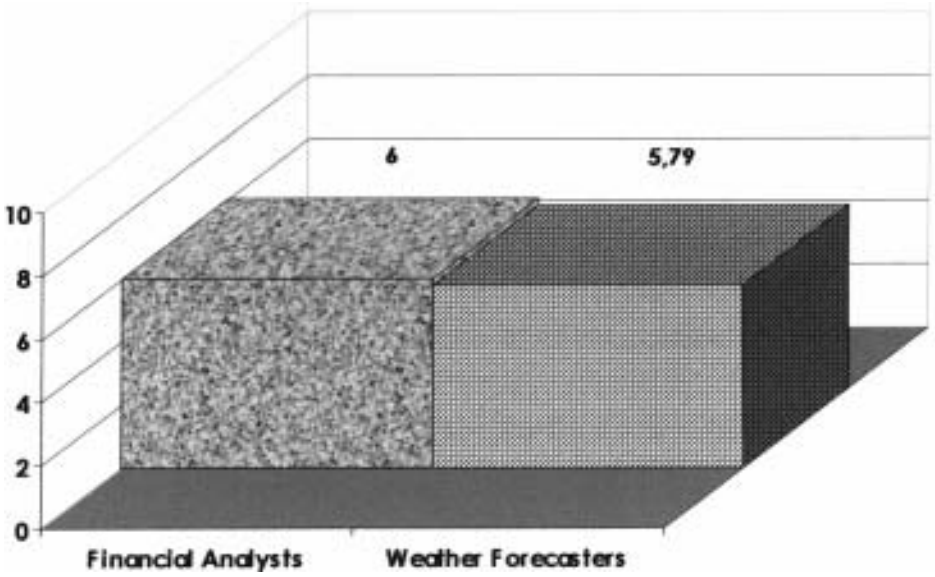
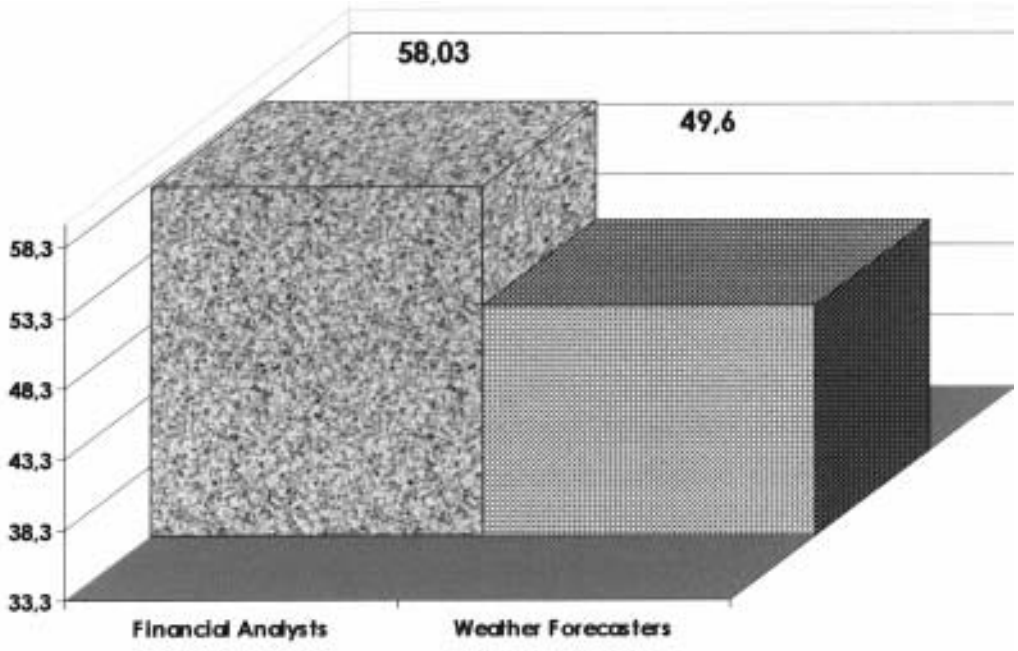



**FIGURE 2**
**Mean Probability Assessments**



overconfidence was significantly higher among financial analysts than among weather forecasters.

When constructing the lists of reasons why a forecast might fail, we included three identical justifications for both groups:

- The events in question are generally unpredictable.
- Common opinions and the behavior of others made me change my opinion.
- In a single prediction there is always a chance of being wrong.

The remaining reasons were more specific to each group, although some were analogous. For the analysis they were classified into several groups:

- I failed to take into consideration factors of various natures.
- There was an unexpected change in the external situation.
- There was manipulation by sophisticated speculators.
- I personally made a mistake, although the event in question is generally predictable.
- I did not have enough personal experience.
- The data and analyses were insufficient.

Figure 3 shows the importance attached by experts in the two groups to different kinds of justifications of the failure of their predictions. Three types of justification dominate in the financial analysts group: 1) unexpected events occurred that changed the situation, 2) in a single prediction there is always a chance of being wrong, and 3) the events in question are generally unpredictable. In the weather forecasters group, the four following justifications dominated: 1) there is no certainty the events can be accurately predicted, 2) in a single prediction there is always a chance of being wrong, 3) lack of personal experience, and 4) insufficient data.

Note that although both groups of experts attach some importance to the probabilistic argument— that there is no certainty that the event in question can be accurately predicted—the weather forecasters attached significantly higher importance to this argument than the financial analysts ($t = 3.6449$, $p = 0.001$). This supports Hypothesis 1.

We performed the following ANOVA for self-evaluations: 2 (Group) × 2 (Time). As Figure 4 shows, a significant interaction effect was found. This shows that after considering why a forecast might fail, weather forecasters, but not financial analysts, changed the assessment of their ability to predict the events in question. This supports Hypothesis 2.

Figures 5 and 6 show the results of 2 (Group) × 2 (Time) ANOVAs for standard deviations of each expert's subjective probability distribution over alternative predictions, and for the assessments of the most probable alternative. In both cases a significant group effect was found. Mean standard deviations of each expert's subjective probability distribution (and the values of the assessments of the most probable alternative) are higher for the financial analysts than for the weather forecasters. Both measures indicate that the weather forecasters show a smaller overconfidence effect than the financial analysts, which supports Hypothesis 3.

## Discussion

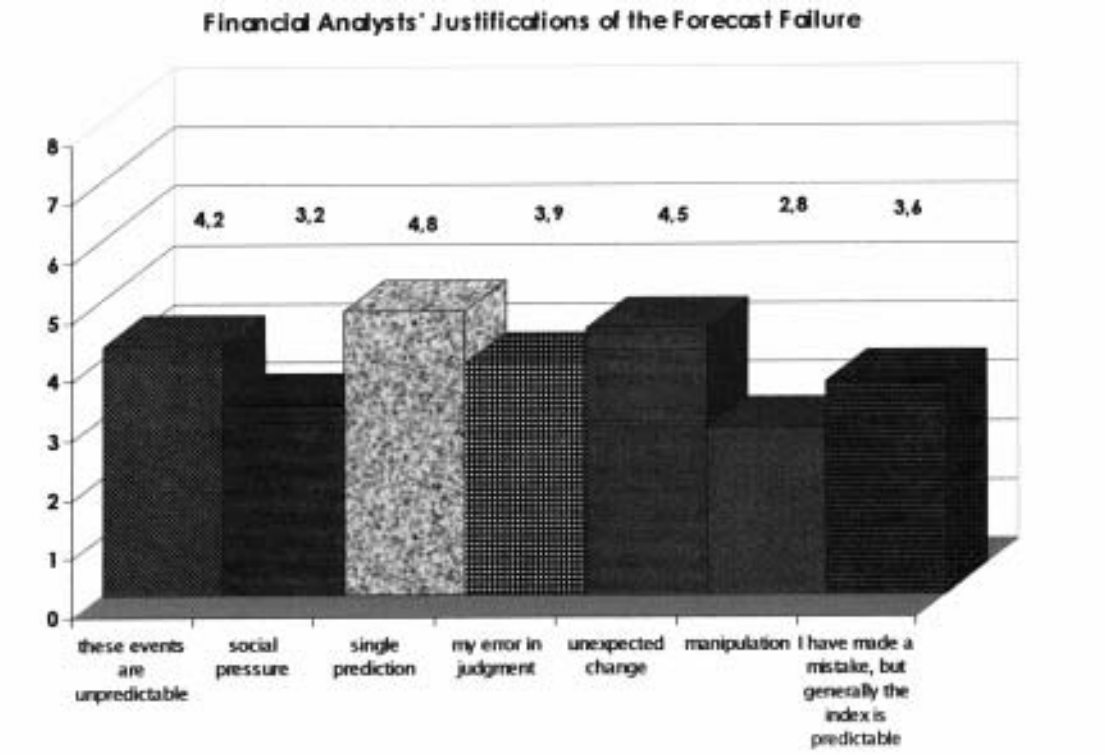
Our research supports the three hypotheses. We found that, in accounting for inaccurate judgments, the two types of experts use various and different arguments to defend their self-esteem. This is in agreement with Tetlock [1999], who described several strategies used by experts to defend their self-esteem. It is possible that such motivation, if strong enough, could prevent experts from learning from experience.

We also found that weather forecasters, i.e., experts in the domain where statistical procedures are used, tend to attach the highest importance to probability, believing that the events in question are generally unpredictable and there is no certainty that they can be accurately predicted. We believe this is attributable to the fact that weather forecasters have to deal with a world that is remarkably complex, but that can be described by some probabilistic relationships. Because seasons repeat cyclically, weather forecasters deal with events of a periodic nature. This world is partially predictable, either through data-based climatological models, or through theory-based NWP models.

Weather forecasters are aware that they are working with a gross approximation of the underlying system and that there will always be a certain amount of uncertainty. Presumably, the same awareness of uncertainty causes these experts to manifest a lower overconfidence effect than experts from the other domain. Indeed, as observed in the U.S. by Murphy and Winkler [1977], meteorologists were exceptionally well-calibrated in the sense that their confidence level was comparable to the actual accuracy of their predictions.

Financial analysts, on the other hand, have to deal with a world that seems completely unpredictable, where even weak probabilistic tendencies are rarely observed. Most of the modern financial theories presume that stock prices follow an approximate random walk pattern (Cootner, 1964; Samuelson, 1965; and Malkiel, 1996). But there is some important evidence that stock price movements deviate from randomness (Lo and MacKinlay, 1999 and Shleifer, 2000), although this has limited meaning for forecasting. In such an arena, no analytical forecasting formula can be

**FIGURE 3**
**Importance Attached to Justifications of the Forced Failure**



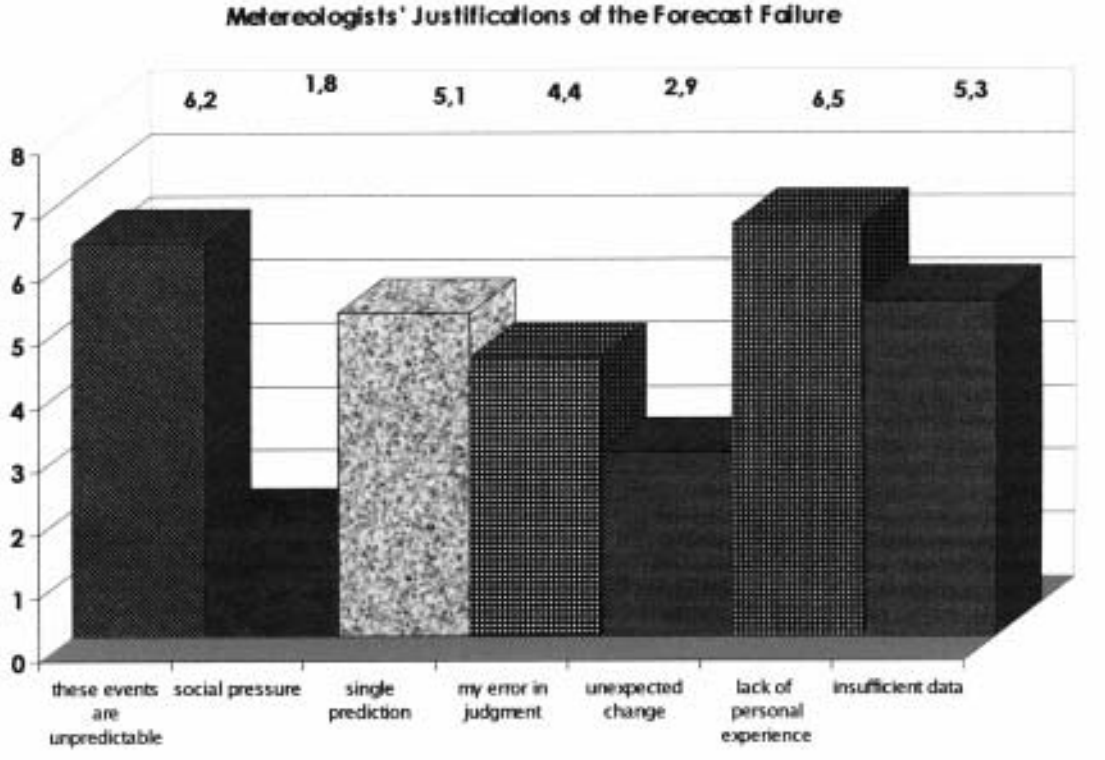
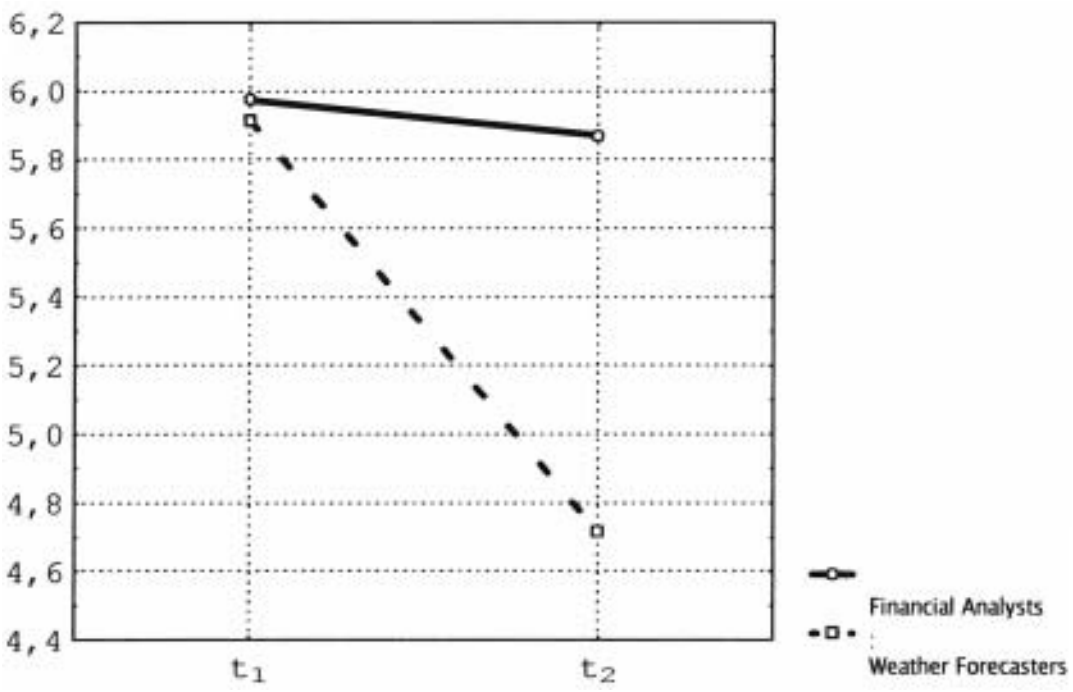
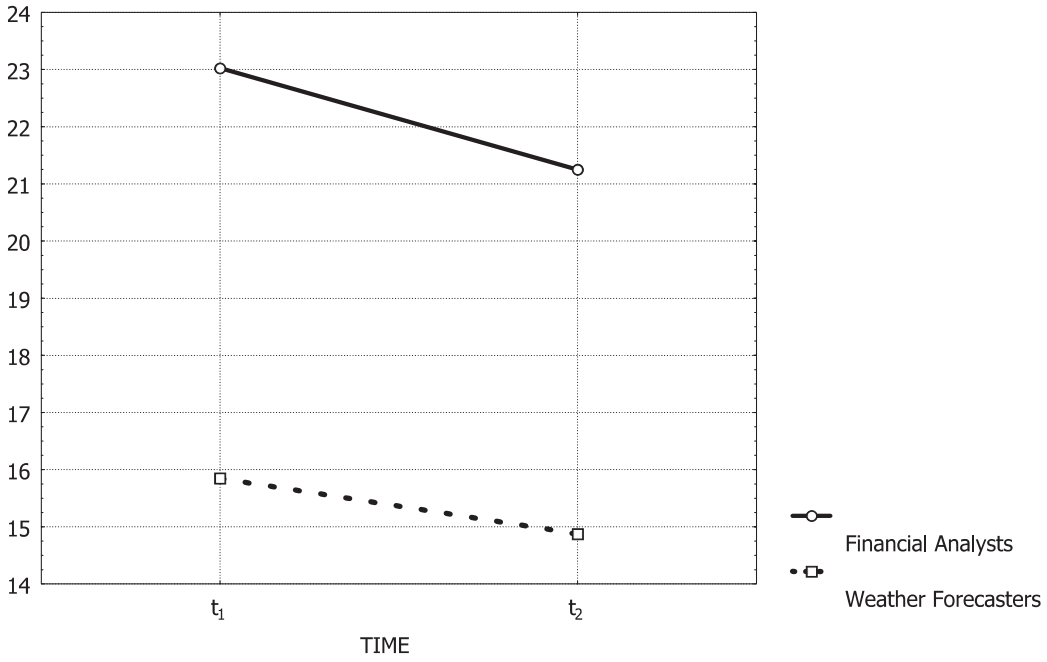



**FIGURE 4**
**Mean Self-Evaluations of Two Groups of Experts for Two Sessions**



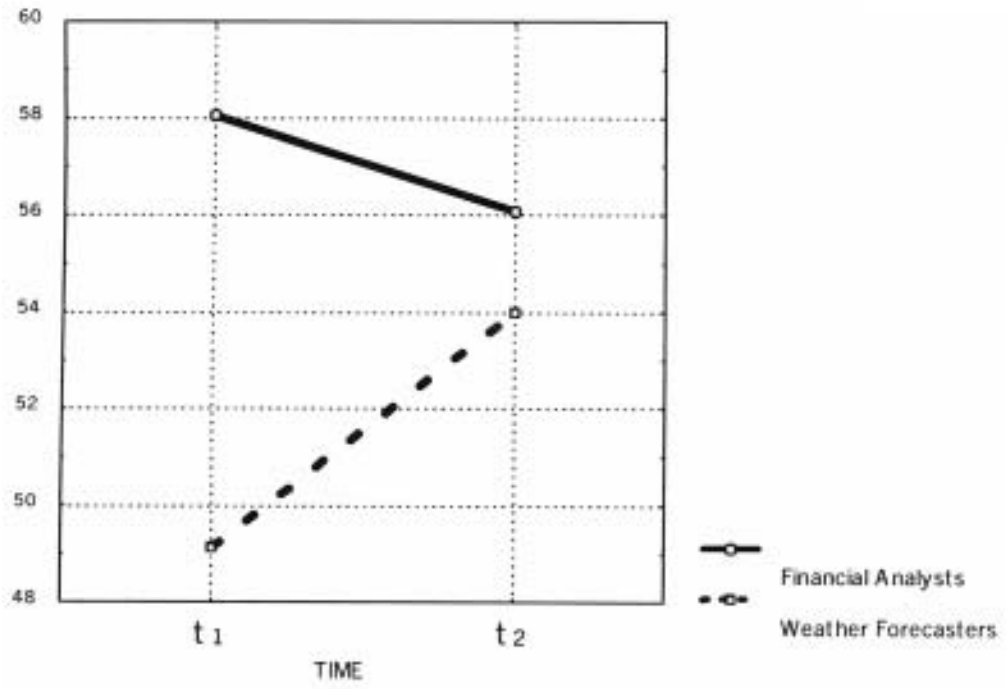

*Note:* ANOVA (Group × Time) Group Effect: n.s. Time Effect: $F = 5.755455$, $p < 0.05$. Interaction Effect: $F = 4.044312$, $p < 0.05$.

**FIGURE 5**
**Mean Standard Deviations of Each Expert's Subjective Probability Distribution for Two Groups of Experts**




*Note:* ANOVA (Group × Time) Group Effect: $F = 12.98852$, $p < 0.001$. Time Effect: n.s. Interaction Effect: n.s.

**FIGURE 6**
**Mean Assessments of Probability for Two Groups of Experts**




*Note:* ANOVA (Group × Time) Group Effect: F = 12.98852, p <0.001. Time Effect: n.s. Interaction Effect: n.s.

used. The situation resembles a casino, where a gambler cannot rely on any regularities to help in making accurate predictions. Therefore, both financial analysts and gamblers have to look to other kinds of heuristics for their predictions.

These heuristics can be based on many factors, for example, any kind of regularity in a previous series of events. In the case of gamblers, there is a well-known fallacy effect, wherein people expect that after a series of, say, nine "tails," a coin should show "heads"). But since these types of accidental indices of predictions are usually based on "natural" human cognitive biases, it is logical that such predictions are formulated with a relatively high certainty. Paradoxically, financial analysts, who actually have less precise knowledge of the underlying system than weather forecasters, can be more self-assured.

The argument discussed so far has been cognitive, but to complete the discussion we must also consider motivational factors. It is possible that certain features of their respective professions may lead to financial analysts being more self-assured than weather forecasters. Note again that the financial analysts expressed a higher level of overconfidence, and, unlike the weather forecasters, their self-evaluations did not decrease after considering why their forecasts might have failed. Thus, the financial analysts behaved as if they had to demonstrate that they did have the *ability* to perform perfect forecasts.

Professor Raymond Dacey from Idaho University suggested that this may stem from the clients of each set of experts. Unless there are severe storms in the area (hurricanes, tornadoes, possible floods), most people who listen to weather forecasts are happy if the forecast is not hopelessly inaccurate. But people who listen to financial analysts (i.e., investors) demand much greater accuracy. If the reality of the predictions is even *slightly* below their expectations, investors can be *very* unhappy. Therefore, in order not to lose their clients, financial analysts must be quite sensitive about their reputations and better skilled than weather forecasters in formulating excuses for their errors. Further research would be needed to confirm these claims, however.

## References

Brehmer, B. "In One Word: Not From Experience." *Acta Psychologica*, 45, (1980), pp. 223–241.

Cootner, P.H. The Random Character of Stock Market Prices. Cambridge: The M.I.T. Press, 1964.

Dawes, R.M., D. Faust, and P.E. Meehl. "Clinical Versus Actuarial Judgment." *Science*, 243, (1989), pp. 1668-1674.

Korn, D.J., and C. Laird. "Hearts, Minds and Money." *Finance Planning*, 28, (1999).

Lichtenstein, S., B. Fischhoff, and L.D. Philips. "Calibration of Probabilities: The State of the Art." In H. Jungermann, and G. de Zeeuw, eds., *Decision Making and Change in Human Affairs*. Amsterdam: D. Reidel, 1977.

Lo, A., and A.C. MacKinlay. A Non-Random Walk Down Wall Street. Princeton: Princeton University Press, 1999.

Malkiel, B.G. A Random Walk Down Wall Street. New York: W.W. Norton & Company, 1996.

Murphy, A.H., and R.L. Winkler. "Can Weather Forecasters Formulate Reliable Forecasts of Precipitation and Temperature?" *National Weather Digest*, 2, (1977), pp. 2–9, 40.

Samuelson, P.A. "Proof That Properly Anticipated Prices Fluctuate Randomly." *Industrial Management Review*, 6, (1965), pp. 41–49.

Sawyer, J. "Measurement and Prediction, Clinical and Statistical." *Psychological Bulletin*, Vol. 66, No. 3, (1966), pp. 178–200.

Shanteau, J. "Expert Judgment and Financial Decision Making." In B. Green, ed., *Risk Behaviour and Risk Management*. Proceedings of The First International Stockholm Seminar on Risk Behaviour and Risk Management, June 12–14, 1995.

Shleifer, A. Inefficient Markets. Oxford: Oxford University Press, 2000.

Tesser, A., and J. Campbell. "Self-Definition and Self-Evaluation Maintenance." In J. Suls and A.G. Greenwald, eds., *Psychological Perspectives on the Self*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

Tetlock, P.E. "Theory-Driven Reasoning About Possible Pasts and Probable Futures: Are We Prisoners of our Preconceptions?" *American Journal of Political Science*, 43, (1999), pp. 335–366.

Trope, Y. "Self-Assessment in Achievement Behavior." In J. Suls and A.G. Greenwald, eds., *Psychological Perspectives on the Self*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

## Notes

1. Naturally, we are aware that the task required clinical judgments from the weather forecasters, which they are not accustomed to providing. Actually, both groups of experts were examined away from their offices and deprived of their usual tools. These artificial conditions were necessary in order to make the experimental results comparable.