

## **Expert Judgments:**

### **Financial Analysts vs. Weather Forecasters**

Tadeusz Tyszka & Piotr Zielonka

Centre for Market Psychology  
of Leon Kozminski Academy of Entrepreneurship and Management  
Warszawa (Poland)

#### **Abstract**

Two groups of experts, financial analysts and weather forecasters - were asked to predict corresponding events (the value of the Stock Exchange Index and the average temperature of the next month). While accounting for inaccurate judgments, weather forecasters attach more importance to the probabilistic nature of the events predicted than financial analysts. Although both groups of experts revealed the overconfidence effect, this effect was significantly higher among financial analysts than among the weather forecasters. These results are discussed from the perspective of learning from experience.

Key words: experts, judgments, forecasts

## **INTRODUCTION**

### **Expert Judgments**

The central question of this paper is what importance experts attach to various justifications of their forecasting failures. In particular the paper focuses on justifications (evoking) odwołujące się to the probabilistic nature of forecast events. We attempt to find out what the consequences for experts' self-confidence are when they are aware of the probabilistic nature of events forecast. Finally, we want to learn how the above is influenced by the type of method used to make expert judgments.

For obvious reasons research on experts' judgments focuses on the question of accuracy of their forecasts. Although accuracy of forecasts is what justifies asking experts for their opinions, much research show that, at least in some domains, experts are not doing particularly well. For example, in recent research by Tetlock (in press) on predictions of

political events, the author found that politics *experts were only slightly more accurate than one would expect from chance.*

Since in many cases a direct evaluation of the accuracy of experts' judgments is difficult or impossible, researchers often study the stability of expert judgments and 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, inter-judge correlations for stock brokers and clinical psychologists were below 0.40, and for auditors this was around 0.70. This indicates that experts in some domains are doing relatively better than in others.

Studies on experts' judgments also show that similarly to lay people, experts are subject to various biases. In particular, they are subject to 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 his/her true responses. As Lichtenstein, Fischhoff and Philips (1982) showed, typically it is 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, Tetlock (in press) in his study on political experts, showed that, although his experts only sporadically exceeded chance predictive accuracy, they assigned extremely high confidence estimates to their predictions. Moreover, the experts were much more overconfident than the customarily observed effect in confidence-calibration research. Experts who assigned confidence estimates of 80% or higher were correct only in 45% of cases. As quoted by Korn & Laird (1999), a high level of overconfidence seems to be characteristic for finance analysts as well and most probably for many other kind of experts.

An important question arises as to how experts react to the confirmation or disconfirmation of their forecasts. This is a crucial question, because the way the experts react to confirmation or disconfirmation of their judgments may determine their learning from experience. And it is well known that the very definition of an *expert* includes the idea of using one's experience. Indeed, according to Webster's (1979) dictionary, *expert* means *having, involving, or displaying special skill or knowledge derived from training or experience.* Do really experts learn from experience? In many questions this is undoubtedly the case. But when it comes to forecasting and learning probabilistic relationships between events the issue becomes open to discussion. One question concerns cognitive abilities to learn probabilistic relationships. Brehmer (1980), who studied people's ability to learn from

experience in a probabilistic situation, concluded that people tend to perceive the 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 forecasts concerns, without doubt, an important issue for the individual – his/her reputation or self-esteem. As research on self-assessment shows, the information on one's own performance activates important motivation to maintain positive self-evaluation (Tesser Cambell, 1983). People *simply want to believe that they are competent, they want others to believe that they are competent; and they will behave so as to maintain a positive self-evaluation* (cf. Tesser Cambell, 1983, p. 5). Of course, the question of motivation is not as simple as this, as it appears that under some conditions people are also interested in obtaining an accurate self-assessment (Trobe, 1983). They are even able to risk damage to their self-esteem in order to obtain an accurate assessment of their abilities.

Correspondingly to the two kinds of motivation, one can expect two different attitudes in experts when they receive the information about the failure of their forecasts. One attitude is to defend one's own self-esteem. Actually, Tetlock in his research on political experts, illustrated several strategies used by experts to defend their self-esteem. For example, one of the most popular defenses 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 predicted events. For example, a weather forecaster whose forecast failed can start to look for causes of errors in his/her predictions – e.g. related to the probabilistic nature of the events predicted. It is interesting to determine when each of these attitudes will dominate.

In connection with these considerations 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-evaluation – do experts whose predictions were confirmed vs. 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?

These questions are referred to in the present paper.

### **Clinical vs. 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. A 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 relations between the events of interest. Indeed, experts predicting events under conditions of uncertainty in some domains (meteorology, econometrics) use special statistical formulae (algorithms), while experts in other domains (stock brokers, clinical psychologists, political scientists) employ what Dawes et al. (1989) refer to as clinical judgments.

Good examples of the two methods are weather forecasting and forecasting in finance. What tools of forecasting are available to a weather forecaster? There is a variety of techniques. Of these two are the most fundamental. One is the climatology method. This involves weather statistics, accumulated over many years in making the forecast. No theory is needed as the predictions are made based exclusively on data analysis. The climatology method only works properly when the weather pattern is similar to that expected for the chosen time of the year. Another is Numerical Weather Prediction (NWP) which uses theoretical forecast models, run on computers, and provides predictions of 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 a mathematical formula.

By contrast, in forecasting stock prices, a financial analyst implements one or a combination of two methods of market analysis - fundamental analysis and technical analysis – and eventually, combining all this information in his or her head it is he/she who makes the prediction (best guess) of stock prices in the future.

It is natural to think that experts in the domains where statistical procedures are used have a better opportunity than those from domains where clinical judgments are used to assess the statistical nature of the events predicted. This should be especially true in such domains as, for example, weather forecasting, where several alternative statistical models of forecasts are available. Such experts may become aware that one might be more or less successful in repeated predictions, but in a single prediction of an event there is always a chance of being wrong. Presumably, this type of experience is less likely in domains where clinical judgments

are used. In these areas few systematic observations of empirical relations between events of interest are carried out. What can be the consequences of such varied experiences?

First, one can think that in their accounting for their failures to forecast accurately, experts in domains where statistical procedures are used may tend to resort to the probabilistic nature of the events predicted more often than experts from domains where clinical judgments are used.

Correspondingly, we formulated Hypothesis 1:

*Experts in domains where statistical procedures are used, while accounting for inaccurate judgments, should attach more importance to the probabilistic nature of the events predicted, and, therefore, – to probabilistic arguments, than experts from domains where clinical judgments are used.*

Second, one can think that activating the process of thinking about the reasons why a forecast might fail – including the probabilistic nature of the events predicted – 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.

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, in agreement with the well-established overconfidence effect, we should expect this effect to occur in all groups of experts. However, experts in domains where statistical procedures are used, being more aware of the probabilistic nature of the events predicted than experts from domains where clinical judgments are used - should be generally less overconfident than the other group.

Correspondingly, we formulated Hypothesis 3:

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

## **METHOD**

Respondents. Two groups of experts were used: financial analysts and weather forecasters. We assumed that the forecasts of the first group are mainly based on clinical judgments, while the forecasts of the second group are typically based on statistical models. (Actually, one of the weather forecasters told us that in this group there is a saying that high-quality weather forecasts should be made with the curtains drawn.).

Predicted events. In the case of financial analysts the forecast concerned the value of the Warsaw Stock Exchange Index (WIG) at the end of 2000 (about one and a 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. In the case of financial analysts there were three intervals of the WIG index – below 15,500; <from 15,500 to 17,000>; and above 17,000 – arbitrarily created to approximate equal-probability intervals. In the case of weather forecasters there were three intervals of temperature – below 7, <from 7 to 9>, and above 9 degrees Centigrade – each of which included 33.3% of the average temperature in April in the last 50 years<sup>1</sup>.

Procedure. Two sessions were run. During session one, respondents were asked to assess on a 9 point Likert scale *how confident they were in the correctness of their assessment of the underlying forces that shape the WIG index or the average temperature in April* respectively. Then, they were asked to rank order the three exclusive and exhaustive events (concerning the WIG index or the average temperature) from the most to the least likely (ties were also 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 up to 100%).

After about two months (when the specified forecasting interval had elapsed), we contacted the respondents again for the session two. First, each respondent was told whether his/her forecast had turned out to be right or wrong. Regardless of whether his/her forecast was right or not, a respondent was presented with a list of reasons why a forecast might fail. These lists were not identical for both groups, but shared some possible reasons why a forecast might fail. (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 7-point scales the importance of each of the reasons (scales were anchored at 1 by “not important at all” and 7 by “very important”).

Then, respondents were given another questionnaire in which they were asked the same questions as in session one, i.e. they were asked to make self-assessment as experts and assign a subjective probability to each of three new events, in the case of financial analysts

---

<sup>1</sup> Naturally, we are aware of the fact that the task required clinical judgements from the weather forecasters, which they are not accustomed to. Actually, both groups of experts, were examined away from their offices and

the value of the WIG index in mid June 2001, and in the case of weather forecasters, 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**

In the group of financial analysts 1/3 of the participants succeeded in their forecast and in the group of weather forecasters approximately 2/3 of the participants succeeded in their forecast. The difference is due to the fact that weather forecasters frequently marked two or three intervals as equally probable. We assumed that someone who had marked two intervals and the predicted value had belonged to one of them had succeeded. The same procedure was implemented for someone who marked all three intervals as equally probable.

None of the phenomena described below are dependent on the accuracy of prediction.

The first two exhibits show the results of session one: (1) how experts evaluated their knowledge of the underlying forces that shape the event in question (self-evaluation) and (2) subjective probability assigned to the most probable alternative. As can be seen from Exhibit 1 mean self-evaluations in the two groups of experts were almost identical and equaled 6 on a 9-point scale. Thus, both groups maintained a rather positive self-evaluation.

-----  
Insert Exhibit 1 here  
-----

Exhibit 2 indicates that both groups of experts revealed the overconfidence effect. As mentioned above, the alternative events were specified in such a way that each of them approximated a 0.33 chance of occurrence, juxtaposed with this, the confidence of experts in their predictions was in both groups much higher than that. However, in agreement with Hypothesis 3, the overconfidence was significantly higher among the financial analysts than among the weather forecasters.

-----  
Insert Exhibit 2 here  
-----

---

deprived of their usual tools. These artificial conditions were necessary in order to make the experimental results comparable.

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

- 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 of them were analogous. For the analysis they were classified into several groups:

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

-----

Insert Exhibit 3 here

-----

Exhibit 3 shows the importance attached by experts in the two groups to different kinds of justifications of the failure of their predictions. As can be seen, three kinds of justifications dominate in the finance analysts group: *unexpected events occurred that changed the situation, in a single prediction there is always a chance of being wrong, the events in question are generally unpredictable and therefore is no certainty that the events can be accurately predicted*. In the weather forecasters group the following justifications dominate: *there is no certainty that the events can be accurately predicted, in a single prediction there is always a chance of being wrong, lack of personal experience, and insufficient data*. One should also notice that although both groups of experts attached some importance to the probabilistic argument – that *there is no certainty that the event in question can be accurately predicted* – the weather forecasters attached to this argument significantly higher importance than the finance analysts ( $t=3,6449$ ,  $p=0,001$ ) This supports Hypothesis 1.

2 (Group)  $\times$  2 (Time) for self-evaluations ANOVA was performed. As can be seen from Exhibit 4, a significant interaction effect was found. It shows that after thinking of the reasons why a forecast might fail is activated, the weather forecasters, but not the finance

analysts change the assessment of their ability to predict the events in question. This supports Hypothesis 2.

-----  
Insert Exhibit 4 here  
-----

Exhibits 5 and 6 show the results of 2 (Group) x 2 (Time) ANOVA's for standard deviations of each expert's subjective probability distribution over alternative predictions, and for the assessments of the most probable alternative. As can be seen, 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 finance analysts than for the weather forecasters. Both measures indicate that the weather forecasters reveal a smaller overconfidence effect than the finance analysts. This supports Hypothesis 3.

-----  
Insert Exhibit 5 here  
-----

-----  
Insert Exhibit 6 here  
-----

**DISCUSSION**

Our research supports the three hypotheses formulated. We found that while 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 (in press) who described several strategies used by experts to defend their self-esteem. One could think 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 high importance to the probabilistic arguments saying that the event in question is generally unpredictable and therefore is no certainty that the event can be accurately predicted. We believe that this sensitivity to probabilistic arguments is due to the fact that the weather forecasters have to deal with a world which is remarkably

complex, but which can be described by some probabilistic relationships. Weather forecasters deal with the events of a periodic nature: seasons repeat cyclically. This world is partially predictable – either through data-based climatological models or through theory based Numerical Weather Prediction models. The weather forecasters are aware that they are working with a gross approximation of the underlying system and that in such an area the uncertainty must be taken into consideration. Presumably, the same awareness of uncertainty causes these experts to manifest a lower overconfidence effect than experts from the other domain. Indeed, as it was 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 which seems to be completely unpredictable, where even weak probabilistic tendencies are rarely observed. Most of the modern financial theories presume that stock prices approximately follow random walk pattern (Cootner, 1964; Samuelson, 1965; Malkiel, 1996). There is some important evidence that stock prices' movements deviate from randomness (Lo, 1999; Shleifer, 2000) but this has a limited meaning for the practice of forecasting. In such an area no analytical formula of forecasting can be used. The situation resembles a casino where a gambler cannot rely on any regularities to help him/her make accurate predictions. Therefore both financial analysts and gamblers try to look for some other kind of heuristics for their predictions. These heuristics can be based on factors of various nature, for example a kind of regularity in previous series of events. In the case of gamblers there is a well known gambler's fallacy effect, wherein people expect that after a series of, say, nine 'tails' a coin should show 'heads'). Since these type of accidental indices of predictions is usually based on "natural" human cognitive biases, it is comprehensible that such predictions are formulated with a relatively high certainty. Paradoxically, financial analysts, having less precise knowledge than the weather forecasters about the underlying system, can be more self-assured.

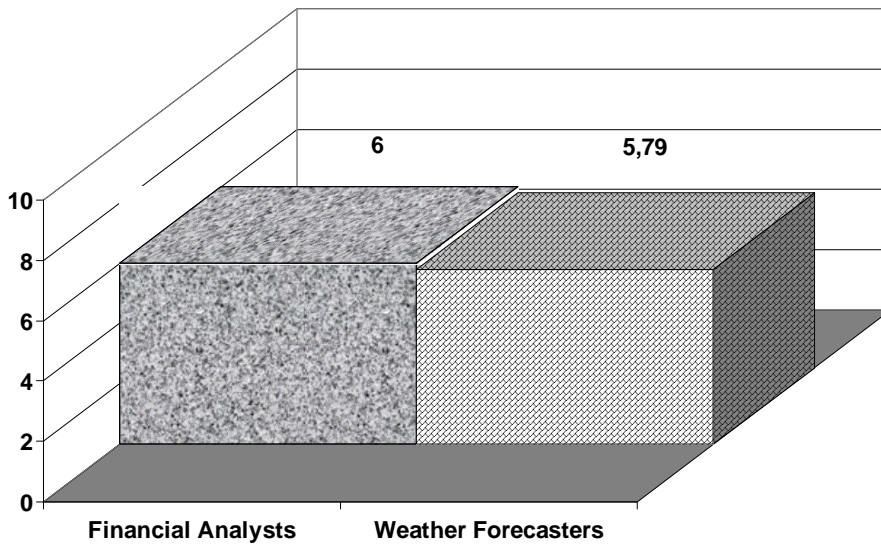
The argument discussed so far has been a of cognitive nature. However, the differences observed between these two groups of experts can also be accounted for by motivational factors. There are perhaps, some features of these two professions that make financial analysts manifest a higher level of self-assurance than the weather forecasters. We found that the financial analysts not only expressed a higher level of overconfidence in relation to the weather forecasters, but also, in contrast to the weather forecasters, they did not decrease their self-evaluation after being motivated to think about reasons why a forecast might have failed. Thus, the financial analysts behaved as if they had to demonstrate the

ability of a perfect forecast of the events in question. Professor Raymond Dacey from Idaho University suggested that this mechanism can pertain to the clients. 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. People who listen to financial analysts (i.e., investors) are very unhappy when the forecasts are only slightly inaccurate, i.e., when the reality is slightly below their expectations. Therefore, in order not to lose their clients financial analysts are very sensitive about their reputation and better skilled than weather forecasters in formulating excuses for their errors. Certainly, further research is needed to confirm these claims.

### LITERATURE

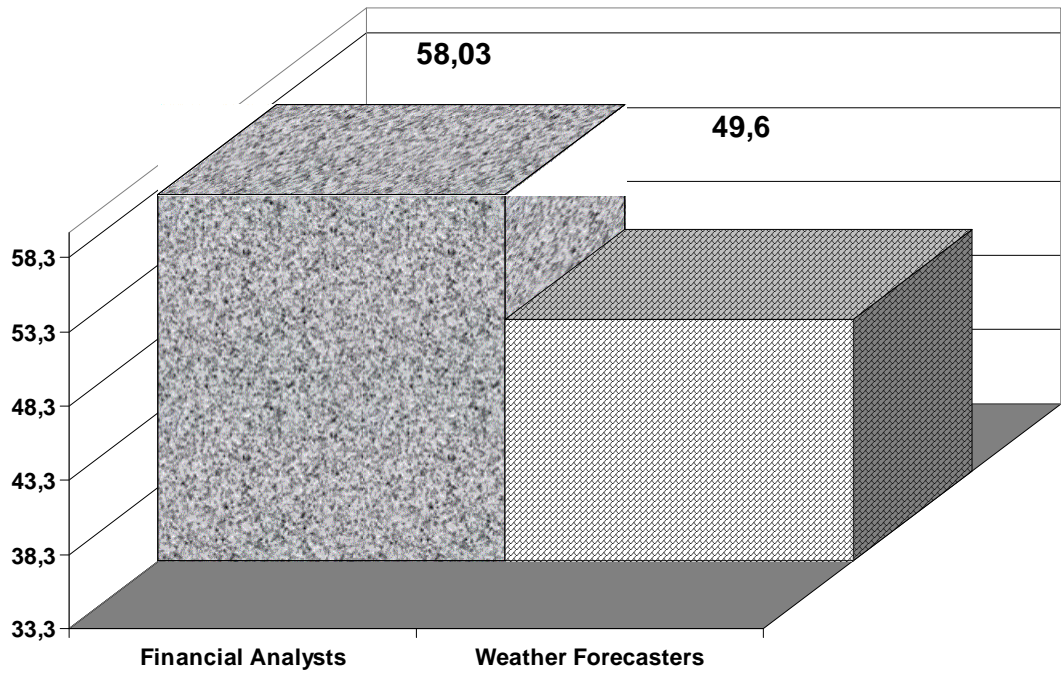
- Brehmer, B. (1980). In one word: not from experience. *Acta Psychologica*, 45, 223-241.
- Cootner, P.H., (1964), *The Random Character of Stock Market Prices*, The M.I.T. Press, Cambridge,
- Dawes, R. M., Faust, D., Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668-1674.
- Korn, D.J., Laird, C. (1999). Hearts, minds & money. *Finance Planning*, 28.
- Lichtenstein, S., Fischhoff, B., Phillips, L.D. (1977). Calibration of probabilities: The state of the art. In: Jungermann, H., & de Zeeuw, G. (eds.), *Decision Making And Change In Human Affairs*. Amsterdam: D. Reidel.
- Lo, A., MacKinlay, A.C., (1999), *A Non-Random Walk Down Wall Street*, Princeton University Press,
- Malkiel, B. G., (1996), *A Random Walk Down Wall Street*, WW Norton & Company,
- Murphy, A.H. & Winkler, R.L. (1977) Can weather forecasters formulate reliable forecasts of precipitation and temperature? *National weather Digest*, 2, pp.2-9, (40),
- Samuelson, P. A., (1965), Proof That Properly Anticipated Prices Fluctuate Randomly, *Industrial Management Review* 6, pp. 41-49, Spring
- Sawyer, J. (1966). Measurement and Prediction, *Clinical and Statistical, Psychological Bulletin*, Vol 66, No 3, 178-200,

- Shanteau, J. (1995). Expert judgment and financial decision making. In: Green, B (ed.), *Risk Behaviour and Risk Management*. Proceedings of The First International Stockholm Seminar on Risk Behaviour and Risk Management, June 12-14,
- Shleifer, A., (2000), *Inefficient Markets*, Oxford University Press,
- Tesser, A. & Campbell, J. (1983). Self-definition and self-evaluation maintenance. In: Suls, J. & Greenwald, A.G. (eds.), *Psychological Perspectives on the Self*. New Jersey: LEA
- Trope, Y. (1983). Self-assessment in achievement behavior. In: Suls, J. & Greenwald, A.G. (eds.), *Psychological Perspectives on the Self*. New Jersey: LEA
- Tetlock, P. (in press). Theory-driven reasoning about plausible pasts and probable futures in world politics: Are we prisoners of our preconceptions? *American Journal of Political Science*.



Test t: n.s

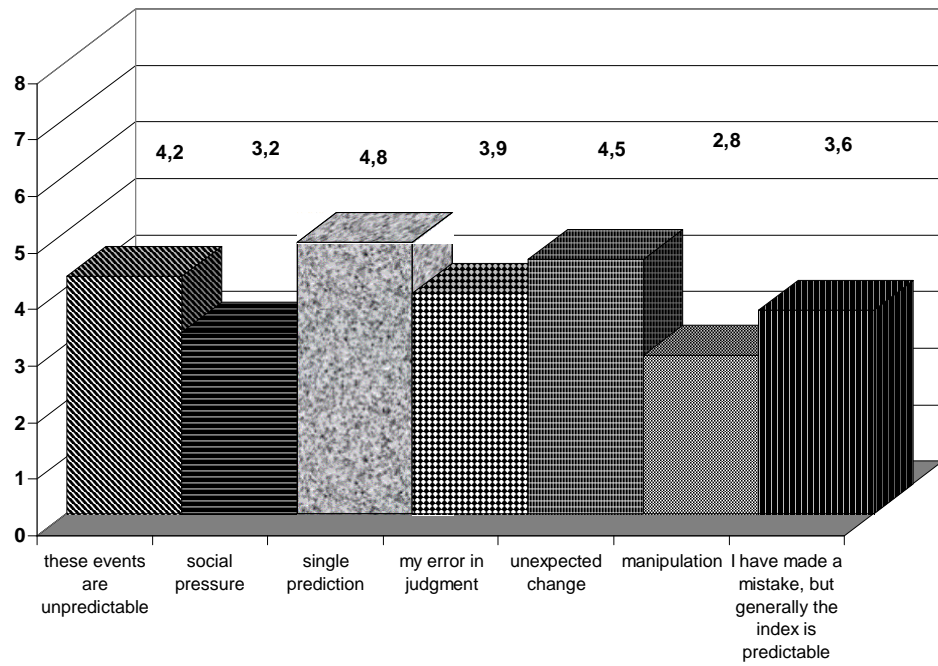
**Exhibit 1. Mean self-evaluations of two groups of experts for session one.**



Test t:  $t=3,299731$ ,  $p<0.05$

**Exhibit 2. Mean probability assessments.**

### Financial Analysts' Justifications of the Forecast Failure



### Meteoreologists' Justifications of the Forecast Failure

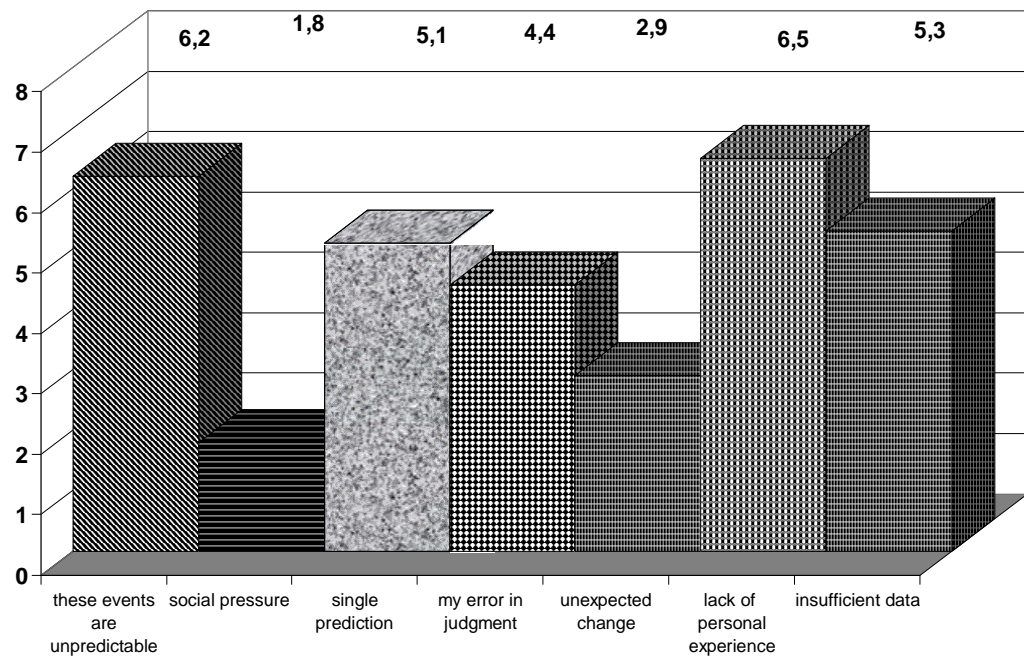
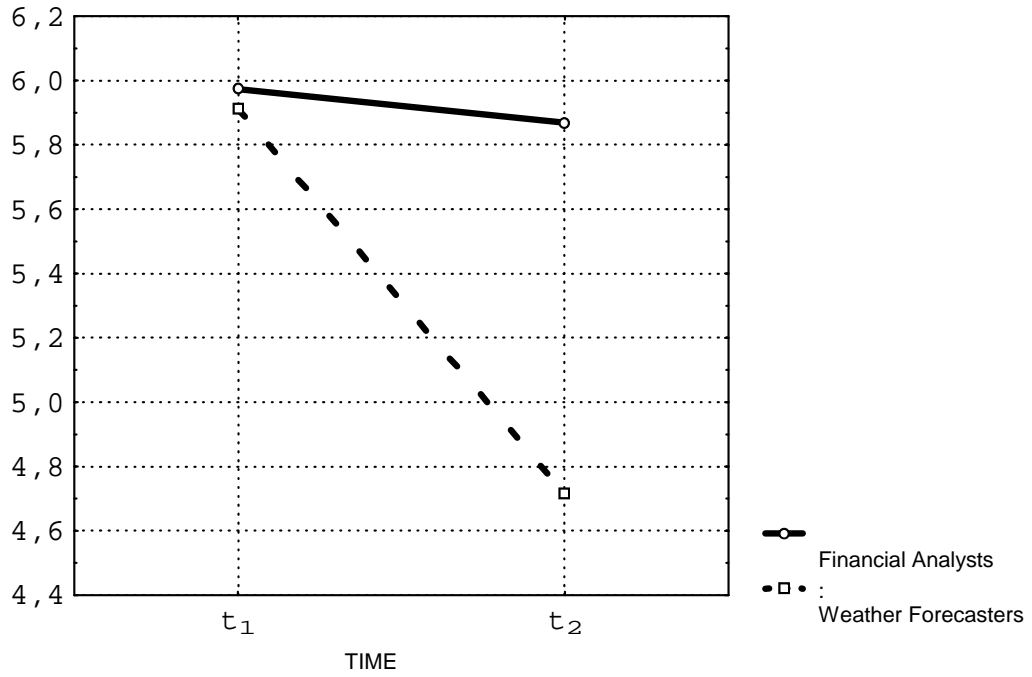


Exhibit 3. Importance attached to justifications of the forced failure.



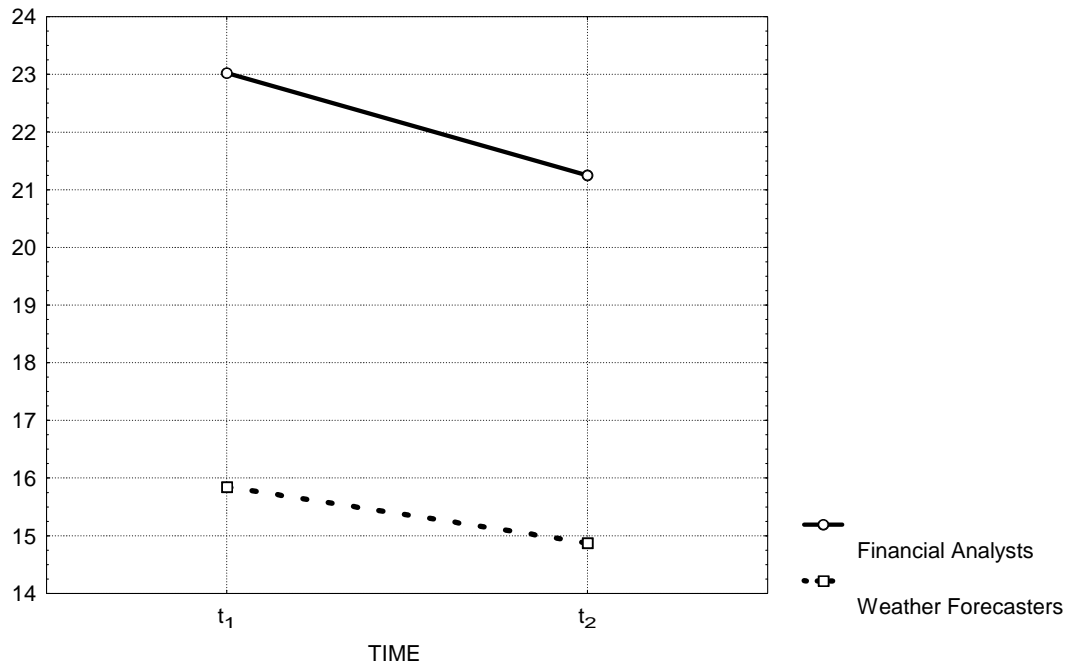
ANOVA (Group x Time)

Group effect: n.s.

Time effect:  $F=5,755455$ ,  $p<0.05$

Interaction effect:  $F=4,044312$ ,  $p<0.05$

**Exhibit 4. Mean self-evaluations of two groups of experts for two sessions.**



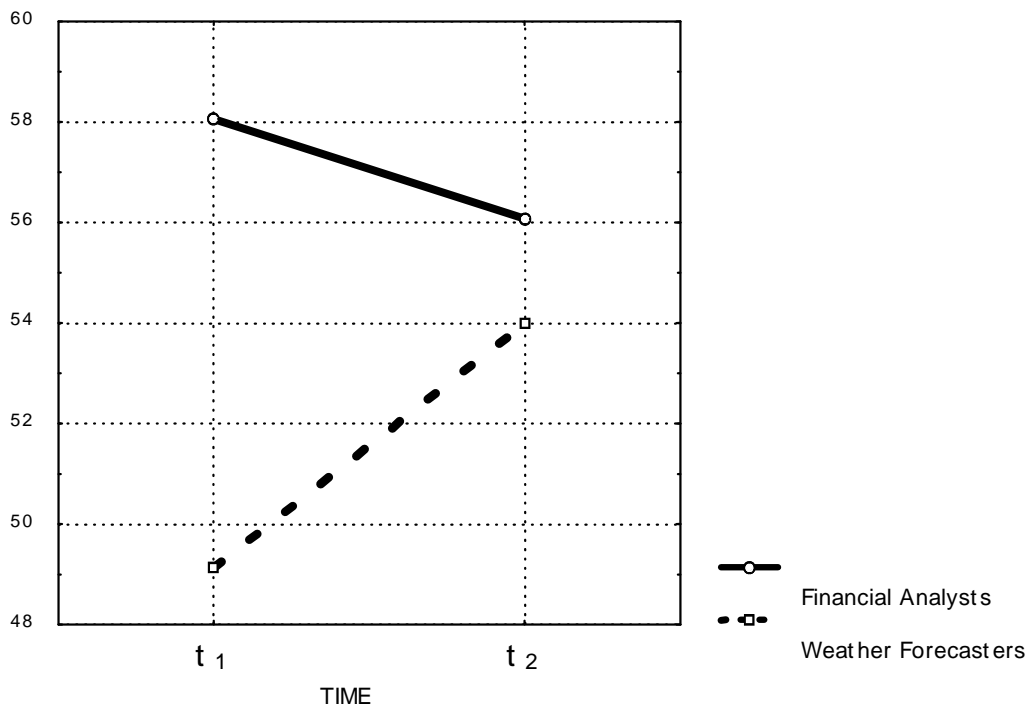
ANOVA (Group x Time)

Group effect:  $F= 12,98852, p<0.001$

Time effect: n.s.

Interaction effect: n.s.

**Exhibit 5. Mean standard deviations of each expert's subjective probability distribution for two groups of experts.**



ANOVA (Group x Time)  
 Group effect:  $F= 12,98852, p<0.001$   
 Time effect: n.s.  
 Interaction effect: n.s.

**Exhibit 6. Mean assessments of probability for two groups of experts.**