Technological Forecasting & Social Change 77 (2010) 355–368

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

The limits of forecasting methods in anticipating rare events

Paul Goodwin [a,*], George Wright [b,1]

[a] School of Management, University of Bath, Bath, BA2 7AY, UK
[b] Durham Business School, University of Durham, Mill Hill lane, Durham City, DH1 3lB, UK

ARTICLE INFO

Article history:
Received 24 March 2009
Received in revised form 17 October 2009
Accepted 20 October 2009

Keywords:
Rare events
Delphi
Scenario planning
Expert judgment
Statistical forecasting

ABSTRACT

In this paper we review methods that aim to aid the anticipation of rare, high-impact, events. We evaluate these methods according to their ability to yield well-calibrated probabilities or point forecasts for such events. We first identify six factors that can lead to poor calibration and then examine how successful the methods are in mitigating these factors. We demonstrate that all the extant forecasting methods — including the use of expert judgment, statistical forecasting, Delphi and prediction markets — contain fundamental weaknesses. We contrast these methods with a non-forecasting method that is intended to aid planning for the future — scenario planning. We conclude that all the methods are problematic for aiding the anticipation of rare events and that the only remedies are to either (i) to provide protection for the organization against the occurrence of negatively-valenced events whilst allowing the organization to benefit from the occurrence of positively-valenced events, or (ii) to provide conditions to challenge one's own thinking — and hence improve anticipation. We outline how components of devil's advocacy and dialectical inquiry can be combined with Delphi and scenario planning to enhance anticipation of rare events.

1. Introduction: what do we mean by predictability?

It is not hard to identify events that have a large impact on the lives of many people, but which were unexpected by most people. Some of these events are natural disasters and some have human causes. Consider the global financial meltdown of 2008. Once such an event occurs it often seems to have been inevitable, with hindsight. Makridakis et al. [1] quote many examples of publicly-available forecasts made by key figures in finance and economics just before the global financial crisis of 2008. These before-event forecasts can now be seen to have been completely wrong. Consider, also, newspaper coverage of terrorist attacks in the US homeland. Such coverage was mute before the 9/11 attacks but, post-event, analysis of the causes took many column-inches. How good are (i) human judgment and (ii) statistical forecasting at anticipating the occurrence of such events? Can techniques that incorporate human judgment in a structured way improve anticipation over and above holistic judgement? This paper analyses these issues and seeks to identify the limitations on our ability to accurately anticipate the occurrence of rare, high-impact, events. We also consider what the implications of these limitations are for organizational planning.

2. The nature of predictability

Assume that all forecasts can ultimately be represented as an objective or subjective probability distribution. We may, of course, only report the forecast in terms of the event we consider most probable (e.g. "I forecast that a Democrat will win in 2012")

* Corresponding author. Tel.: +44 1225 383594; fax: +44 1225 826473.
E-mail addresses: mnspg@management.bath.ac.uk (P. Goodwin), george.wright@durham.ac.uk (G. Wright).
[1] Tel.: +44 191 33 45427; fax: +44 191 33 45201.

0040-1625/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.techfore.2009.10.008

or as a measure of central tendency of the distribution, such as the mean (e.g. “The expected level of demand next year is 2500 units”). Also, we are excluding forecasts that may be expressed in fuzzy terms (e.g. “I think that there's a good chance that the economy will perform well”). Based on this assumption, Wright and Goodwin [2] argue that the term, predictability, can be interpreted on two levels. First, predictability can relate to the capability of forecasters to produce a well-calibrated probability distribution. Perfect calibration would be achieved, for example, if it rains on 10% of days when we have said that the probability of rain is 10%. If it rains on more, or less, than 10% of those days then our probability assessment is mis-calibrated — i.e., we may be over-confident or under-confident in our assessment. Similarly, if our forecast is simply reported as the mean of the distribution, we would expect the mean outcome in the future to be close to this value if our forecast is well-calibrated. Of course, when forecasts are made relatively infrequently we may not able to measure calibration.

Alternatively, even if forecasts are made frequently, we are by definition unlikely to be able to collect much data for occasions when rare events occur; so measuring our capability of assigning appropriate probabilities to such event will be problematical. Nevertheless, the concept is still useful as a criterion for explaining what we mean by poor quality, or high-quality, forecasts. Note than when we use the term predictability in this sense we are referring to the capability of carrying out the prediction task in a valid way. Such predictability can be high even when there are a large number of possible events that can occur, each with a low probability of occurrence, as long as our estimates of these probabilities are well-calibrated. In a draw in the UK national lottery one of 13,983,816 different sets of numbers can be selected. However, such events are predictable, in this first sense, because we can determine a perfectly calibrated probability for each set.

Second, predictability can be interpreted as relating to the dispersion of the probability distribution — the more dispersed this is then the higher will be the expected error associated with a particular quantitative point forecast or, if the forecast is expressed as a statement that a particular event will occur, the lower will be the probability that the statement will be correct. For example, by this, second definition of predictability, if the future demand for a product is approximately normally distributed with a standard deviation of 250 units, then the outcome of future demand will be more predictable than if the standard deviation had been 500 units. Thus, in this second sense, the predictability of sets of numbers in the UK national lottery is low. If you predict that a given set of numbers will be drawn then you only have a 1/13,983,816 probability of being correct.

Wright and Goodwin [2] argue that if well-calibrated probabilities can be obtained, decision theory can be used to make rational decisions on the basis of them, even if the predictability (in the second sense) is low [3]. If we have a reliable probability estimate for the occurrence of an earthquake in a particular county in the next ten years we can use a rational process to assign an appropriate level of resources in anticipation of that event, even if this probability is very low. If we do not have a reliable estimate then we may assign an inappropriate level of resources. It is therefore the first form of predictability — the ability to establish well-calibrated forecasts — that is the topic of this paper (hereafter we will use the term in this sense only). While our prime interest is in events that have the potential to have a major impact it should be noted that the probability of these events is not necessarily low — the probability of an important event may actually be quite high, we may simply have not recognised this. We will first examine the potential reasons why predictability in a given situation may be low. Then we will compare the effectiveness of methods that are designed either to improve predictability — or to allow for effective planning when predictability cannot be improved. We then consider the implications for planning in organizations.

3. Six causes of low predictability

3.1. Sparsity of reference class

Predictability will be greater when we have data on a large set of similar events (i.e., a large reference class) from which relative frequency information can be obtained. This will be the case when events are defined more generally — the greater the specificity of the definition, the smaller will be its reference class. The number of terrorist attacks of any kind in the world in the course of a year is therefore more predictable than the number of terrorist kidnappings occurring in the course of a week. Large reference classes are akin to larger samples of a population — they allow us to make more reliable assessments of the underlying probability distribution. Large reference classes also lend themselves to statistical analysis, so that judgmental biases can be avoided in the estimation task. Thus for some events, like the number of earthquakes or hurricanes that might occur in the world in the course of a year, it is possible to establish relative-frequency-based, objective, probabilities and these probabilities are likely to be well-calibrated. In contrast, novel events, for which there are no past analogies, such as “the Gulf stream will stop flowing within twenty years” are likely to be highly unpredictable.

3.2. Reference class is outdated or does not contain extreme events

Reference classes are bound to be incomplete because they are samples — and they are samples only the past and not the future. This will be a problem when systems that impact on the occurrence of those events are subject to fundamental changes. Reference classes are also likely to be biased because very rare events with potentially massive impacts are, by definition, unlikely to be included in the sampling — so that their probabilities of occurrence (or even possibilities) are discounted due to sampling bias. The use of a reference class can, therefore lead to poorly-calibrated forecasts for the occurrence of rare, high-impact, events.

3.3. Use of inappropriate statistical models

Even when a reference class is rich in data, there is a danger that poorly-calibrated forecasts may be obtained because of erroneous assumptions and the use of an inappropriate probability distribution. This may occur when people mistakenly view the reference class as being a reliable sample of possible events and ignore the issues mentioned above. For example, Taleb [4] reports that financial models often assume that changes in stock prices follow a normal distribution yet, the stock market crash of Black Monday represented a fall of 20 standard deviations from the mean and hence, if the normal distribution assumption is true, should only have occurred once in 'every several billion lifetimes of the universe' [4].

Models are also bound to be simplifications of real systems and may not fully account for complex interactions between elements of these systems. This is likely to be true of models of the economy, weather systems or the human body [5]. The effect of minor changes in one part of the system or in initial conditions can be amplified through these interactions. Thus the range of uncertainty indicated by the model may under estimate the true range so that the generated probabilities are poorly calibrated. Drawing analogies from systems biology, Orrel and McSharry [5] have suggested that using a single model to capture the behaviour of these complex systems is inappropriate and that what is needed is the use of different approaches to model different aspects of systems. These multiple approaches will require the collaboration of experts in different fields. However, these suggestions have, as yet, been untested in areas such as climate, economic or political forecasting and the authors themselves appear to have some doubts about their likely success when they argue that "instead of trying to predict the future, [perhaps] we should use models to better understand a system's behaviour."

3.4. The danger of misplaced causality

Most models will be based on the assumption about the causal relationships between variables. However, a coherent theory of causality, which provides a good fit to data in the reference class and which may have the support of a broad consensus of experts in the relevant field, does not establish that the causality exists. For example, there is a strong correlation between carbon dioxide emissions and global temperatures and a coherent theory to explain this linkage which has received widespread scientific support. However, this has not prevented challenges to this theory of the cause of global warming. Correlations may be spurious (e.g. they may result from hidden third factors), or they may only apply in the conditions that are relevant to the reference class data. Moreover, when human judgment is involved (see below), correlations may be illusory [6] with preconceived correlations being confirmed in the judge's mind by the selective recall of instances that accord with the belief in the correlation. The fallacy that a high correlation necessarily implies causation is widely encountered and can be a powerful influence of people's reasoning.

3.5. Cognitive biases

When the reference class contains insufficient cases for statistical estimation, human judgment is often used to estimate the probabilities of events occurring. Much of the research on the quality of human judgment of probability has stemmed from the work of Tversky and Kahneman [7] who argued that people use simple mental strategies or heuristics to cope with the complexities of estimating probabilities. While these heuristics can sometimes provide good estimates and reduce the effort required by the decision maker, they can also lead to systematically biased judgments. The three main heuristics identified are:

i) *Availability*. Here, events within the reference class which are vivid, recent, unusual or highlighted by the media are readily recalled or envisaged and therefore assigned high probabilities. Availability can be a reliable heuristic since frequently occurring events are usually easier to recall so the higher probabilities estimated for them should be reliable. However, the ease with which an event can be recalled or imagined sometimes has no relationship to the true probability of the event occurring. For example, some events are easily recalled precisely because they are unusual and rare. By contrast, events that have never occurred, or only occurred in the distant past, may be assigned a de-facto probability of zero, or near-zero.

ii) *Representativeness*. This heuristic describes a tendency to ignore base-rate frequencies and was demonstrated by Tversky and Kahneman in a series of experiments where participants were asked to judge the probability that an individual had a particular occupation. Participants were given both base rate information and a low-quality, but stereotypical, description of the person. The finding was that valid base-rate information was ignored. This and related studies indicate that even when useful reference class information is salient for utilisation in forecasting it will be ignored in favour of ephemeral, low-validity, individuating information. Indeed, Kahneman and Lovallo [8] have argued that people tend to see each individual forecasting problem as unique when it would best be thought of as an example of the broader reference class of events. Hence they tend to pay particular attention to the distinguishing features of the problem in hand and reject analogies to other instances of the same general type as superficial. For example, Cooper et al. [9] found that entrepreneurs who were interviewed about their chances of business success produced assessments that were unrelated to objective predictors such as college education, prior supervisory experience and initial capital. Moreover, more than 80% of them described their chances of success as 70% or better while the overall survival rate for new businesses is as low as 33%.

Gigerenzer [10] argues that we are simply not equipped to reason about uncertainty by assessing subjective probabilities for unique events but that we can reason successfully about uncertainty with frequencies. For example, the entrepreneurs might have been asked instead: "What percentage of new businesses is successful?" However, obtaining relative-frequency-based assessments is not feasible for rare events because a reference class of previous forecasts or historic frequencies is not available. If human thinking is best expressed, and thought of, as that of frequency thinking rather than probabilistic thinking this conceptualization clearly does not help in the anticipation of high-impact rare events.

iii) *Anchoring and insufficient adjustment*. Here, forecasts that are used in the decision process may be biased by forecasters anchoring on the current value and making insufficient adjustment for the effect of future conditions. Alternatively, there may be a tendency to anchor on the probability of single events when estimating the probability that a particular combination of events will occur. For example, if an individual component of a system has a 0.9 probability of functioning perfectly over a given time scale this probability may unduly influence an estimate of the probability that all 100 components of the system will function perfectly over the period in question.

3.6. Frame blindness

The frame refers to how one views and structures a prediction problem. It involves determining what must be predicted, the form the prediction will take (e.g. point estimate or prediction interval), what factors are likely to impinge on the event that is to be predicted, the consequences of inaccurate prediction, the likely reliability of the prediction and the effort and resources that it is appropriate to devote to the prediction task. Since predictions are made to inform decisions the prediction frame will be closely aligned with the way that the corresponding decision has been framed. Frames are bound to be simplifications of real problems and each of them will only give a partial view of a prediction problem. For example, different frames will emphasize different potential influences on the event that is being predicted or they may attach different degrees of importance to the potential errors associated with the prediction. Difficulties can arise when a single frame is used unquestionably by forecasters, perhaps because of habit or professional specialism. Managers' mental models of the world, exemplified by the use of a single frame, are analogous to single visual perspectives on a scene. One viewpoint through a window frame may mean that only part of the external world is in view while another observer, looking through a different window frame, may see more (or less) of the external environment. Additionally, the past experience of the observer shapes his or her (mis)interpretation of events that occur.

In one study the variability between individual managers' mental models of competitive structures in the UK grocery retailing industry was examined [11]. Considerable variation was found in the nature and complexity of industry views from managers both within and between companies. This diversity was associated with the functional roles that individual managers held. Barr et al. [12] addressed the issue of why some organizations are able to realign their strategy with a changing environment, whilst others are not, offering a cognitive explanation for the lack of organizational renewal. They argued that 'human (cognitive) frailties mean that managers' mental models of the competitive environment may be incomplete or inaccurate, and that these models 'often fail to change in a timely manner in response to a changing environment' (p 17). At the same time, political pressures within the organization act to quell dissonant or 'deviant' opinion, which recognise the true, paradigm-threatening nature of the information (see also [13]).

All of this indicates that habitual frames of reference may come to dominate thinking and changes in the world that may herald the occurrence of rare, high-impact events may not be recognised as such. Consider, for example, the dramatic sub-prime mortgage crisis that started in the US. The causal factors behind the crisis now seem obvious, with hindsight. But these causes seem not to have been so obvious to the finance industry insiders, a-priori.

4. Solutions?

We next assess the extent to which these six causes of the low predictability of high-impact, rare events can mitigated by approaches that have been proposed either to improve the calibration of forecasts or to enable effective planning to take place when it is assumed that unpredictability cannot be reduced. Table 1 provides a summary of how well each of these methods is likely to impact on, or be impacted by, the six causes of unpredictability that we have just outlined. Some of the strengths and limitations of these methods may also apply when the probabilities of frequently occurring methods are being assessed. However, in these circumstances there is more chance of rapid and data-rich feedback on the accuracy of the forecasts so that errors and biases may be recognised and corrected more quickly.

4.1. Statistical forecasting

When there is a large reference class of relevant data, statistical forecasting has the advantage that this data can be handled completely and efficiently, thereby precluding the cognitive biases associated with human judgment (although, in economic forecasting, data series used in model building are often inaccurate and liable to be revised, sometimes after significant delays). That said, there is nothing inherent in statistical forecasting to warn that the forecasting problem may have been inappropriately framed and that attention is being paid to forecasting the wrong phenomena. For example, we may focus our efforts on forecasting the behaviour of competitors in our industry or the effects of our marketing mix, when the real impact on our company's well being will come from new industries based on novel technologies. Events sometimes have a high-impact precisely because they represent a change from events contained in the reference class that is thought to be relevant.

In addition, judgmental and motivational biases may apply in the choice of forecasting method and data. Changing conditions in an industry may mean that data on only the most recent members of the reference class are relevant. The longer the lead time of

Table 1
How the methods relate to the sources of unpredictability.

Source of unpredictability	Statistical forecasting	Expert judgment	Decomposed judgment	Structured analogies	Judgmental adjustment to statistical forecasts	Delphi	Prediction markets	Scenario planning
Sparsity of reference class	Unreliable in these circumstances	May outperform statistical methods in these circumstances	May outperform statistical methods in these circumstances	Supports best use of available cases in reference class	May lead to improvements over statistical forecasts in these circumstances	Addressed, in principle, by exchange of reasons	May outperform statistical methods in these circumstances	Analysis of causal interactions allows participant to see beyond existing reference class
Inappropriate reference class	Unreliable in these circumstances	Expert may focus on explaining current circumstances and lose the wider picture	Unreliable if reference class is used	Unreliable in these circumstances	Unreliable if reference class is used	Addressed, in principle, by exchange of reasons	Fast response to new information may counter this problem	Danger of anchoring scenarios in current economic conditions and current media concerns
Inappropriate statistical model	Unreliable in these circumstances	Not applicable	Not applicable	Not applicable	Mixed evidence on whether adjustments can compensate for model	Not applicable	Not applicable	Not applicable
Misplaced causality	Model may embed false assumptions about causality	Expert may focus on ‘pet theory’ and defend its use with vigour	Decomposition structure may emphasize false causality assumptions	Selection of analogies may be predicated on false causal assumptions	Adjustments may reflect illusory correlations	Reasons exchanged may reflect particular theories about causality which may be false	Participants in the market may be influenced by the paradigm which is currently popular.	Scenarios depend on beliefs that particular causal chains apply
Cognitive biases	Avoids problem for a given data set, but biases may apply in selection of data and method	Experts are likely to suffer from cognitive biases	Biases should be reduced by decomposition	Biases in recall of similar cases should be reduced	Unreliable in these circumstances, but structured methods may improve reliability	Addressed, in principle, by dialectical nature of process and averaging of individual estimates	Mitigated in part by aggregation of individual estimates	Simulation heuristic may lead to inappropriate confidence that a detailed scenario will unfold
Frame blindness	Not addressed	Unlikely to be addressed by expert associated with a particular ‘school of thought’	Not addressed. Structure of decomposition will be predicated on current frame	Not addressed	Not addressed	Addressed, in part, by dialectical nature of process	No mechanism inherent in the method for challenging this	May reinforce existing frame unless ‘remarkable people’ are employed to challenge participants' frames of reference

the forecast then the greater the danger that few, if any members, of the reference class will be useful. Alternatively, in the short term, there may be a tendency to fit models only to recent members because of over-reactions to events and perceptions of changes that are really only noise [14].

Causality can also be problematical for statistical methods. Correlations between variables do not prove causality so the selection of independent variables needs to be based on some external theory of what is likely to account for variations in the dependent variables. However, such theories are themselves likely to have been informed or supported by the extent to which they fit past observations and they may be inapplicable to conditions that will apply in the future.

Assumptions that forecasting errors are normally distributed may be tempting because they simplify the analysis and allow access to a well-established body of techniques. Such assumptions have been prevalent in portfolio analysis variance at risk (VaR) techniques [15]. However, these methods tend to underestimate the probability of extreme events when the 'true' distribution is 'fat-tailed'. This is because most of the data used in fitting the model is close to the central tendency of the distribution and data on extremes is by definition, rare. Extreme value theory has attempted to avoid this bias by concentrating analysis on the extremes and using distributions such as the generalised extreme value distribution (GEV). This has a tail index parameter which determines the thickness of a distribution's tail. However, extreme value theory still faces a number of challenges [15]. The small amount of data that is available on extremes has to be used to determine if the 'true' distribution is indeed fat-tailed and to estimate the parameters of the distribution. In addition, defining a threshold for what is deemed to be an extreme event, and hence should be included in the estimation process, can be problematical. Lowering the threshold increases the number of observations available for estimating the tail index so that the estimate is more precise, but it is also likely to bring in observations that are closer to the central tendency so that the bias in the estimate is increased.

The complete absence of extreme events, above a certain threshold, from the reference class means that they can only be forecast by extrapolation. Extrapolation involves the strong assumption that the relationship between dependent and independent variables remains the same beyond the observed data set. For example, assumptions of linear relationships may not apply far beyond the observed data.

4.2. Expert judgment

Statistical methods will be unreliable when membership of an appropriate class is sparse. In these cases recourse may be made to the use of experts' judgmental forecasts. Research on the quality of calibration performance of experts' probability assessments — usually with respect to forecasting performance — has been found, in several instances, to be very good; for example, [16] (financial interest rates); [17] (horse racing); [18] (the card game, Bridge); and, most strikingly, in weather forecasting [19]. Conversely, in several instances poor calibration has been found — for example [20] (clinical psychologists) and [21] (maize judges). More recently, Tetlock [22] collected 82,361 political and economic forecasts from experts asking them to estimate probabilities for various events. They performed worse than chance.

Judgmental probability forecasts are routinely generated in weather forecasting. Indeed, the official forecasts issued by the National Weather Service in the United States are subjective probability forecasts. Murphy and Brown [19] evaluated these subjective forecasts and found that, for certain categories of weather, they were more accurate than the available objective statistical techniques. The weather forecasters have a very large amount of information available, including the output from statistical techniques. They also receive detailed feedback and have the opportunity to gain experience of making forecasts under a wide range of meteorological conditions. Furthermore, they have considerable practice in quantifying their internal state of uncertainty. These circumstances may well be ideal for the relatively successful application of judgmental as compared to purely quantitative forecasting.

More widely, Bolger and Wright [23] and Rowe and Wright [24] have argued that in many real world tasks, apparent expertise (as indicated by, for example, status) may have little relationship to any real judgment skill at the task in question. In Bolger and Wright's review of studies of expert judgmental performance they found that only six had showed "good" performance by experts, while nine had shown poor performance. Bolger and Wright analyzed and then interpreted this pattern of performance in terms of the "ecological validity" and "learnability" of the tasks that were posed to the experts. By "ecological validity" is meant the degree to which the experts were required to make judgments inside the domain of their professional experience and/or express their judgments in familiar metrics. By "learnability" is meant the degree to which it is possible for good judgment to be learned in the task domain. That is, if objective data and models and/or reliable and usable feedback are unavailable, then it may not be possible for a judge in that domain to improve his or her performance significantly with experience. In such cases, Bolger and Wright argued, the performance of novices and "experts" is likely to be equivalent and they concluded that expert performance will be largely a function of the interaction between the dimensions of ecological validity and learnability — if both are high then good performance will be manifested, but if one or both are low then performance will be poor.

Wright et al. [26] studied expert life-underwriters and attempted to ensure that the expert-task match was as strong as possible (given experimental limitations), and that ecological validity was high, and yet still obtained expert performance that was not much better than lay person performance. This result suggests that the underwriting task is not truly "learnable", i.e., it is not one for which there is regular feedback on the correctness or otherwise of judgments. Indeed, in the training of underwriters, performance is assessed according to the similarity of junior underwriters' judgments to those of their seniors [27]. Once "trained," underwriters receive infrequent performance-related, objective feedback about the correctness of their judgments, and indeed it would be difficult to provide such feedback, given that a "poor" judgment might turn out to be insuring an applicant who subsequently died of a condition after perhaps 20 years of a 25-year policy.

As such, the tasks performed by *other* professional risk assessors may also be un-learnable. For example, in the case of major hazards in the nuclear industry there may be no risk/judgment feedback at all and the calibration of expert judgment cannot be assumed. Similarly, recall the validity of expert predictive judgments about the likelihood magnitude of human infection by "mad cow disease" resulting from eating beef from herds infected with Bovine Spongiform Encephalopathy (BSE) in the early 1990s and the subsequent, poorly predicted, mortality rates [25]. Here the fact that the event was novel and unique precluded the availability of feedback. We conclude that the common sense assumption of the veracity of expert judgment of the likelihood of rare, high-impact events is ill-founded. The lack of a reference class of prediction-outcome data for such rare events means that experts cannot learn from feedback, over time. It follows that bias in expert judgment is, likely to be prevalent — since solely heuristic processes can be utilized by experts in the generation of forecasts. In addition, Tetlock [22] found that the experts in his study were skilled at inventing excuses for the errors in their forecasts. (e.g. "I was almost right", "my timing was just off" or "I made the right mistake") This would further reduce any chance they had of learning from the very limited outcome feedback that they received.

4.3. Structured judgmental decomposition

As indicated earlier, judgmental forecasts may be subject to cognitive biases. Decomposition of the forecasting task into smaller and hence easier judgmental tasks, it is argued, should improve the quality of any estimates elicited, including probabilities [28]. For example, the greater ease with which the component tasks can be carried out may reduce reliance on over-simplifying heuristics and hence reduce the effect of their associated biases.

Decomposition, using event trees or fault trees [29] may be particularly helpful when probabilities of very rare events have to be estimated. Availability bias, caused for example by the reporting of unusual events in the media, may lead to the probabilities of very rare events being over estimated while people may also have difficulty in distinguishing between probabilities like 0.00001 and 0.0000001 [3]. In these circumstances, an event tree could be formulated to depict the combinations of events which might foreshadow a rare event. The tree would then allow probability estimates to be made for these pre-cursor events, rather than the rare event itself. Many of these events may be relatively frequent and be associated with large reference classes so that statistical methods could be used to estimate their probabilities. These probabilities can then be multiplied to establish the estimated probability of the rare event.

Decomposition has other potential advantages. In decision analysis, for example, the separation of the probability estimation tasks from the consideration of the attractiveness of outcomes, may reduce the effects of wishful thinking or optimism bias. In addition, the process of explicit quantification 'forces participants to express their assumptions and beliefs, thereby making them transparent and subject to challenge and improvement' [30]. It also can act as an antidote to groupthink [31] where risks are ignored or underplayed by groups of decision makers. By forcing explicit consideration of the possibilities, decomposition may help to bring hitherto unrecognised opportunities or threats to the surface so that appropriate and timely action can be taken.

However, decomposition is not a panacea for the elicitation of judgmental forecasts. The events for to which the decomposition is being applied may depend on a restricted or inappropriate decision frame. As a result the wrong problem may be addressed and probability assessments may not be carried out for events which represent fundamental changes from the status quo and which can have major impacts in the future. Moreover, the structure of the decomposition is likely to depend on particular beliefs about what constitutes the casual chain of events. In addition, there may be problems in motivating forecasters to engage in decomposition because it involves an explicit exposure of one's assumptions, which may then be subject to challenge, while, if the decomposition is detailed it can involve considerable time and effort. Motivation is likely to be particularly adversely affected where the decomposition method is unfamiliar to the person making the forecast or there is scepticism about the technique that is being used to implement it [28].

4.4. Structured analogies

Another approach to improving judgmental forecasts involves drawing the forecaster's attention to what is available in the reference class by highlighting the role of analogies. Without this support people may rely informally on their ability to remember similar cases so availability bias may result from a propensity to recall recent or unusual cases. When the use of judgment is appropriate it is likely that the membership of the reference class will be small. Because of this some researchers have proposed approaches that allow access to the reference class to be structured so that improved inferences can be drawn from it despite its sparsity. For example, Lee et al. [32] investigated ways of improving judgmental estimates of the effect of future sales promotions by providing a database of past promotions and deploying an algorithm which displayed the promotions that were most similar to the forthcoming promotion, together with their estimated effects on demand. Obtaining a useful number of analogies necessarily involved selecting past promotions which differed to some extent from the target promotion (e.g. in their timing, type or sales region) so Lee et al.'s forecasting support system also provided a simple facility that allowed the user to explore the likely effect of these differences. This helped them to estimate the size of adaptations they needed to make to the promotion effects of the selected cases when making their forecasts. The structured use of analogies has also been investigated in the context of conflict forecasting by Green and Armstrong [33]. Here, experts were asked to recall conflicts that were similar to the target case, to state the outcome of these conflicts and to rate their degree of similarity with the target. An administrator then combined this information to produce a forecast. In both of these studies the structured approach to the use of reference class information led to significant improvements in forecast accuracy.

However, in the case of rare events, there is a danger that this emphasis on past analogies may distract the forecaster's attention away from the possibility of events which are not within the existing reference class, particularly rare and extreme events — a situation which is highly likely when membership of the reference class is sparse. Also, the selection of similar events through either algorithms or expert judgment also may be predicated on a particular view of causality (e.g. that the effects of sales promotions are dependent on the characteristics that have been selected for storage in a database or that conflicts that are judged to be similar on a set of characteristics will be resolved in the same way because of these characteristics).

4.5. Statistical forecasting with judgmental intervention or adjustment

Some systems manifest regular behaviour that is occasionally disturbed by the effects of foreseeable special events. For example, a time series of the demand for a product may exhibit regular seasonal patterns, which are disturbed when the product is promoted or subject to a change in taxation. In this situation statistical methods are likely to provide well-calibrated forecasts during normal periods. However, the effects of the special events (these are sometimes referred to as "broken leg" cues) may be relatively unpredictable. When these events are infrequent or unique, the absence of a large reference class of similar events will preclude the effective use of statistical methods. In these cases forecasters may apply their judgment to estimate the effects of the special event.

In companies, managers commonly adjust statistical baseline forecasts to take the effect of special events into account, while economists often apply judgment to the components of econometric models [34,35]. Laboratory and real world studies have demonstrated that such adjustments typically improve the accuracy of the baseline forecasts [36,37]. There is, however, mixed evidence that they can compensate for situations when an inappropriate statistical model has been applied to the data [38,39]. Moreover, research also suggests that there is much scope for enhancing predictability in these situations. First, decisions on when to intervene are often poor with a tendency towards over intervention as people falsely see special cases in random movements in the graph or are motivated to adjust forecasts to reinforce a sense of ownership of the forecasting process [36,40]. Second, estimates of the size of the required adjustment are subject to cognitive biases. As a result, they often poorly calibrated with the outcomes of the special events. Clearly, the decomposition and structured analogies approaches outlined in the last section may be effective in improving judgmental adjustments as well as forecasts that are wholly based on judgment.

4.6. Delphi

Judgment, alone, is used in the Delphi procedure where multiple individuals are initially required to give separate numerical judgments or forecasts — often years into the future and often for high-impact events. These forecasts are, likely to be revised in the light of feedback provided anonymously by other members of the Delphi panel, over a number of subsequent 'rounds' or iterations. Response stability found across panellists, is the signal to cease additional iterations and take the average of the final round as the Delphi yield.

Delphi's effectiveness over comparative procedures, at least in terms of judgmental accuracy, has generally been demonstrated. In a review of empirical studies of Delphi, Rowe and Wright [41] found that Delphi groups outperformed 'statistical' groups (which involve the aggregation of the judgments of non-interacting individuals) in twelve studies, underperformed these in two, and 'tied' in two others, while Delphi outperformed standard interacting groups in five studies, underperformed in one, and 'tied' in two. This trend is all the more impressive given that many laboratory studies of Delphi effectiveness have used simplified versions of the technique (e.g. with limited feedback) in simplified contexts (e.g. using non-expert, student subjects) that might be anticipated to undermine the virtues of the technique.

Although research suggests that Delphi allows improved judgement compared to alternative methods, as demonstrated in these 'technique comparison' studies, the reasons for this are still unclear, given relative dearth of 'process' studies that have attempted to establish the precise mechanism for improvement in Delphi. Generally, it is assumed that Delphi improves judgemental accuracy because of the feedback provided between rounds — in conjunction with the panellists' anonymity. Rowe and Wright [42] compared three feedback conditions: an 'Iteration' condition over rounds without feedback from the members of the Delphi panel), a 'Statistical' feedback condition (involving median values and range of estimates), and a 'Reasons' feedback condition (involving reasons from the Delphi panellists along with their numerical estimates). They found that, although subjects were less inclined to change their forecasts as a result of receiving Reasons feedback than other types, when they did change forecasts, this change tended to be for the better, leading to a reduction in error. Although subjects tended to make greater changes to their forecasts in the Iteration and Statistical conditions than in the Reasons condition, these changes *did not*, in general, improve predictions.

As such, there is indicative evidence that the receipt of reasons why a particular numerical forecast is being advocated by a panel member is a useful source of information that can be used to improve other panellists' predictions. However, note that the focus of the Delphi procedure is on the prediction of single target variables such as the date of occurrence of a future event or a point estimate of an uncertain future quantity. Delphi is a well utilized procedure and most applications focus on forecasts of a 20–25 year horizon. The exchange of reasons between panellists can, in principle, alert panellists to inappropriate framings, biases in the recall of similar cases, utilisation of inappropriate reference classes, cognitive bias, and inappropriate views of causality underpinning the unfolding of event chains. However, much depends on the degree of communication of the reasoning processes underpinning a particular panellists' prediction. In most Delphi applications, many predictions are sought from expert panellists and so, in practice, exchange of elaborated reasons may be attenuated. Also, exchange of reasons has not, to date, been a priority in practice — most applications of Delphi have involved the exchange of numerical estimates only.

4.7. Prediction markets

Prediction markets offer an alternative method of obtaining estimates from groups. Participants trade contracts which typically stipulate that their owner will receive a sum of money (say $1) if a particular event occurs and nothing otherwise. The current price of the contract is taken to be the participants' aggregate view of the probability that the event will occur. Certain theoretical conditions have to be met for this to be the case. (e.g., that traders are risk averse and their beliefs are independently normally distributed around the true value [43,44]) However, the reliability of the approach may be robust to departures from these assumptions and empirical studies of the performance of a diverse range of markets indicate that they do yield accurate results [45]. Prediction markets offer the advantage that they rapidly respond to the latest information which may reduce the danger of heavy dependence on out-of-date members of the reference class Also the aggregation of individual estimates may counter the cognitive biases of individual forecasters.

Nevertheless many of the reports of accurate forecasts obtained from prediction markets relate to circumstances where there was a relatively small set of possible outcomes (e.g. outcomes of research and development projects, winners of Presidential elections, successes of new products, which films will be box-office successes, Oscar winners and outcomes of sports events). There is less evidence about their success in producing well-calibrated probabilities for rare events. Indeed, there would have to be some awareness of the possibility of such an event in the first place in order for a contract relating to it to be formulated. Moreover, the high level of stock markets' prices before the credit crunch of 2008 suggests that markets may not be good predictors of such events. The majority of participants in a market may be influenced by predominant views about causality presented by the mass media. In addition, when anonymous reasons underlying judgments are exchanged in a Delphi process, people have an opportunity to learn and hence improve their estimates. In prediction markets no such information is shared so there are no opportunities to challenge the potential frame blindness of individual participants.

4.8. Scenario planning

The practice of scenario planning implicitly accepts that managers are *not* able to make valid assessments of the likelihood of unique future events and that ‘best guesses’ of what the future may hold may be wrong. This view is in harmony with Gerd Gigerenzer's argument that probability theory does not apply to single events. Advocates of scenario planning also argue that it can counter groupthink by allowing minority opinions about the future to have ‘airtime’, relative to majority opinion.

How do scenarios achieve this? The first point to note is that a scenario is not a forecast of the future. Multiple scenarios are pen-pictures of a range of *plausible* futures. Each individual scenario has an infinitesimal probability of actual occurrence but the *range* of a *set* of individual scenarios can be constructed in such a way as to *bound* the uncertainties that are seen to be inherent in the future — like the edges on the boundaries surrounding a multi-dimensional space.

Scenarios focus on key uncertainties and certainties about the future and use this information to construct pen-pictures in an information-rich way in order to provide vivid descriptions of future worlds. By contrast, subjective probabilities entered into a decision tree provide numerical values that can be used in an expected utility calculation. The judgment process that produced such numbers is often not verbalized or recorded. When individuals disagree about their subjective probabilities for a critical event, then decision analysis practice is often to take an average, or weighted average, rather than to explore, in detail, the reasoning processes underlying individuals' assessments. Inherent in such analysis is the assumption that it is useful and possible to attempt to predict the future, whereas scenario planning assumes that the best that can be done is to identify critical future uncertainties and plan for the range of futures that could, plausibly, unfold. Essentially, scenarios highlight the causal reasoning underlying judgments about the future and give explicit attention to sources of uncertainty *without* trying to turn an uncertainty into a probability. A major focus is *how* the future can evolve from today's point-in-time to the future that has unfolded in the horizon year of the scenario — say 10 years hence. The relationship between the *critical* uncertainties (as they resolve themselves — one way or the other), important pre-determined trends (such as demographics, e.g. the proportion of the US population who are in various age bands in, say, 10 years' time) and the behaviour of actors who have a stake in the particular future (and who will tend to act to preserve and enhance their own interests within that future) are thought through in the *process* of scenario planning such that the resultant pen-pictures are, in fact, seen as plausible to those who have constructed the scenarios.

Fig. 1 gives two examples of such causal analysis using data from a recent intervention, conducted by one of the authors, in a major EU bank involved in residential mortgage lending. The scenario method used was the “intuitive logics” approach — see [46,47] for more detail. The two clusters which were viewed by workshop participants to be both (i) of the highest uncertainty and (ii) the highest impact on the bank's operations are illustrated.

Note that, in general, the two clusters that result from application of the intuitive logics approach to scenario construction will each contain a mix of pre-determined elements and critical uncertainties that are causally linked together. The four scenarios that are constructed at the next step are derived from the resolution of events within each cluster into two major outcomes — with each of the two outcomes of the first cluster then being combined with each of the two outcomes of the second cluster (see [46], chapter

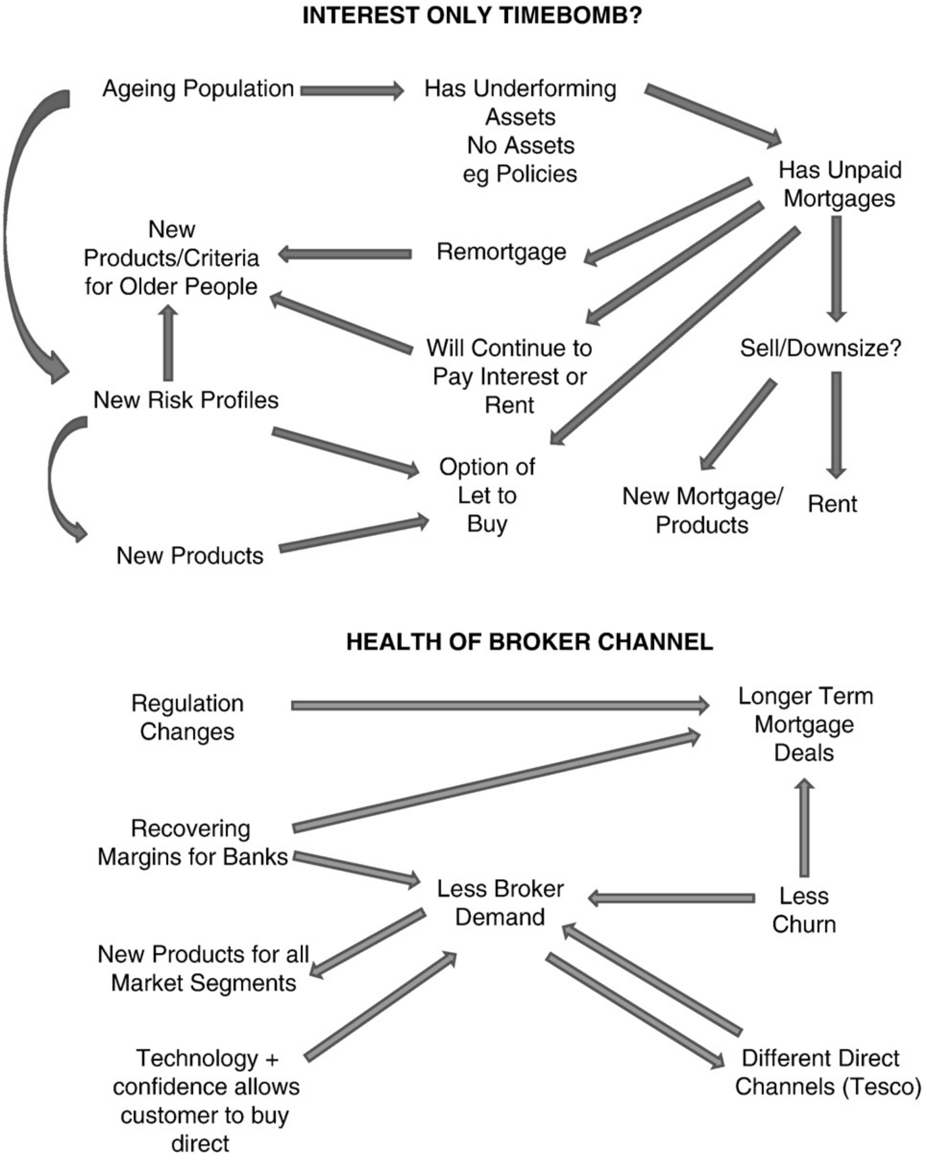
INTEREST ONLY TIMEBOMB?
Ageing Population
Has Underforming Assets No Assets eg Policies
Has Unpaid Mortgages
New Products/Criteria for Older People
Remortgage
Will Continue to Pay Interest or Rent
Sell/Downsize?
New Risk Profiles
Option of Let to Buy
New Mortgage/ Products
Rent
New Products
HEALTH OF BROKER CHANNEL
Regulation Changes
Longer Term Mortgage Deals
Recovering Margins for Banks
Less Broker Demand
Less Churn
New Products for all Market Segments
Technology + confidence allows customer to buy direct
Different Direct Channels (Tesco)

Fig. 1. Two high-impact, high-uncertainty clusters.

7, for more detail). Thus, resolution of the contents of the two high-impact, high-uncertainty, clusters drive the development of the storylines of the four resultant scenarios. The development of the four storylines will, in practice, also utilise other uncertainties and pre-determined elements that have been generated by scenario workshop participants but which are seen, by these participants, to have less impact on the focal issue of concern of actual occurrence.

Note that scenario planning is a practitioner-derived approach to dealing with uncertainty in decision making. It is not based on an axiom system — as is decision analysis — and so different practitioners tend to promote different methodologies to construct scenarios. As we have seen, scenario thinking emphasizes the construction of causal 'storylines' that describe how the future will unfold. Such a way of anticipating the future seems to be quite natural. For example, Willem Wagenaar in a study of how judges reach decisions in courtrooms has found, analogously, that judges and juries do not weigh probabilities that a defendant is guilty 'beyond reasonable doubt'. Instead, such decision makers evaluate scenarios that describe *why* and *how* the accused committed the crime. One such scenario is, in principle, contained in the prosecution's indictment. The prosecution tells the story of what happened and the court decides whether that is a true story or not. 'Good' stories provide a context that gives an easy and natural explanation of why the 'actors' behaved in the way they did. So, storytelling via scenario planning may be a natural way of making sense of the world. Because of its focus on causality, scenario planning is intuitively more attractive to managers and the take-up of scenario planning has been extensive compared to decision analysis — see [47]. Within a scenario planning workshop, decision makers experience and acknowledge the continuing fluidity of an emerging decision context.

Scenario planning does not evaluate options against uncertainties in a single process of analysis. Instead, once the range of plausible futures has been defined, these futures can be utilized over an extended time period as and when new decision options are developed and subsequently tested in the 'windtunnel' conditions.

However, scenario planning is not without problems in aiding the anticipation of rare, high-impact events. Availability bias can enter scenarios, such that recent and current media-emphasized concerns (e.g. of financial downturns) replicate themselves in to-be constructed scenarios. These practice-recognised issues have been labeled as "future myopia". By contrast, as Wright et al. [48] note, one way, used in practice by scenario practitioners, is to provide challenge to the decision makers' mental models by the introduction of "remarkable people" into the strategic conversation — i.e., by including, as participants, in a scenario exercise those individuals (often from outside the host organization) who hold disparate and contradictory views on key uncertainties. In the scenario intervention conducted by the authors with a EU bank's residential mortgage business, described earlier, the participant directors evidenced no recognition of factors that could lead to the — then just months away — sub-prime meltdown in the US and its subsequent impact on the UK housing market. In fact, at the time of our scenario intervention the bank was considering increasing its sub-prime exposure! Whether or not the inclusion of "remarkable people "in, what was, a purely internal scenario planning exercise would have placed the sub-prime meltdown on the scenario agenda is unknown. Scenario planning practitioners argue that between-workshop activity spent on researching the nature of critical uncertainties identified in earlier workshops will also add to the quality of a strategic conversation about the nature of the future — but empirical evidence on the benefit of such desk-based research has also not been conducted.

Interestingly, only one extant study has provided an investigation of the impact of the use of scenario planning on subsequent and contemporaneous corporate performance [49]. In that study, the authors measured the degree of use of scenario planning in both water industry firms and IT consulting firms. However, the achieved questionnaire returns from the firms in these industries was low (22 Water and 25 IT) and so, in our view, even indicative conclusions cannot be drawn. Clearly, as yet, the benefits of scenario planning on an organizational performance have not been empirically demonstrated.

5. Can the anticipation of rare events be improved?

5.1. Protective strategies

The above discussion reveals that all of the extant methods contain weaknesses. Of particular concern are those possible high-impact events that are implicitly assigned a probability of zero. As such, decision makers using any or all of these methods will still be susceptible to surprises that may have severely negative consequences or represent huge missed opportunities. Makridakis and Taleb [50] argue that we should accept that accurate predictions of the occurrence of rare, high-impact events are not possible and so should adopt protective strategies — such as hedging by the use of financial "covered puts". They argue that we should buy insurance to limit the downside of negatively-valenced events (such as a huge loss of a major industrial plant) but allow the unlimited upside of positively-valenced events (such as a possible huge gain by investing a small amount in a speculative venture). Makridakis et al. [1] argue that business strategies should be built to the same analogous standard as buildings that are designed to withstand low-probability, but high-impact, earthquakes. Taleb [51] argues for redundancy in financial investment by retaining "idle" capital — so-called de-leveraging. He notes that human beings have some duplicate organs and also some organs can take on new functions — so-called degeneracy. Thus, maximising redundancy, although increasing costs and restricting the possibility of leveraging resources, enables survival in difficult times. In a similar vein, Wright and Goodwin [2] argued that the decision maker should be alert to the degree to which any major strategic option is: (i) flexible — i.e., investment can be up-scaled or down-scaled at any point in the future; (ii) diversified — i.e., following the option that diversifies the firm's current major offering(s) by providing either a different technology base, a different production base, or a different customer base; (iii) insurable. This prescription can be implemented as a necessary check-list that must be completed in any option evaluation or as part of a more formalised, multi-attribute, evaluation of options against scenarios [52].

5.2. Attempting to prepare for all possible high-impact events

An alternative to having strategies to provide protection against unknown events which are assumed to be completely unpredictable is to try to identify all possible high-impact events that might occur and make contingency plans to deal with them. For example, in the sphere of crisis management, Pearson et al. [53] noted that many organizations prepare for the crisis that they believe most probable or will have most impact if it occurs. These authors argue that, instead, "… the best-prepared organizations compile a crisis portfolio for an assortment of crises that would demand different responses… this may seem a wasteful approach but… the most dangerous crises… cause greater trouble, specifically because no one was thinking about or preparing for them" (p 55). However, the cost–benefit trade-off of preparing an organization for all possible crises is not addressed in the extant literature. Nor is a systematic approach offered to enable managers to rank-order crises for differential attention.

5.3. Widening the range of possible scenarios

In the sphere of scenario planning, Wright and Goodwin [2] argued for an enhancement of the scenario planning process by creating a range of more extreme scenarios than those that result from the use of the intuitive logics scenario development methodology, described earlier. Wright and Goodwin argued that scenarios should encompass a wider range of uncertainties in order to anticipate rare high-impact events. For example, a conventional range of scenarios for the UK economy may contain GDP growth figures ranging from −2% to +5%. But how secure can decision makers be that this represents the complete range of possibilities? Rather than moving forward through causal chains to arrive at scenarios, as in conventional, intuitive logics, scenario planning, Wright and Goodwin's alternative is to work backwards from an organization's objectives. Here, the ranges of possible achievement (worst possible and best possible case) for each of the main objectives can be extended (i.e., made more extreme) and decision makers can be asked whether there they can envisage particular interactions of pre-cursor events that make these, more extreme, best- and worst-case levels of achievement plausible. Note the word "plausible" — plausibility implies that the causal events underpinning a major scenario outcome can be articulated. As such, outcomes such as "GDP growth of 3000%" would not be deemed plausible and so would not be part of any set of such extreme scenarios. In a similar vein to Wright and Goodwin's backward logic scenario method, Makridakis et al. [1] argue that strategic thinkers should create a "virtual time-machine" and imagine that rare, high-impact, events have, in fact, happened. Next the strategists should attempt to think-through their causation. In short, methods for widening the range of constructed scenarios are now under development but, of course, the resultant scenarios may still not contain the particular rare events that actually occur. This is especially true if the causal unfolding of these events is, a-priori, opaque to scenario workshop participants.

5.4. Practical proposals for enhancing Delphi and scenario planning by incorporating devil's advocacy and dialectical inquiry

Schweiger et al. [54] observe that discussion and other interaction amongst top executives are the common ways in which information is shared and evaluated. But groups of decision makers often smooth over conflict and the social pressure for social harmony amongst individual group members is strong — such that group members become more concerned with retaining the approval of fellow members than coming up with good solutions to the task in hand. As Janis [55] noted, in discussing his concept of groupthink, these processes can lead to the suppression of ideas that are critical of the decision on which the majority of a group is converging and, as such, there can be a failure to examine the risks of preferred decisions, a failure to re-appraise initially-rejected alternatives, a failure to work out contingency plans, and an increasing feeling of invulnerability in the group's decision. As a remedy, Janis argued that the leader should (i) withhold his own opinion — since in hierarchical organizations subordinates will also tend not to criticise opinions proffered by those who are higher up the hierarchy, (ii) encourage new ideas and criticisms, (iii) make sure that the group listens to minority views, and (iv) use processes designed to delay forming an early consensus.

Of the methods that we have reviewed to aid the anticipation of rare, high-impact events, only Delphi and scenario planning provide some degree of argument-based challenge to thinking. As we have seen, Delphi does this by the anonymous dialectical exchange of arguments for particular points of view. By contrast, scenario planning does this by engaging a process whereby detailed causal stories for alternative plausible futures are constructed. But as we argued and demonstrated, conventional scenario planning may, in fact, replicate and reinforce existing frames of the future unless "remarkable people" are employed to challenge these framings. Groups tend to share information that the individuals have in common and the probability that a piece of information is shared in group discussion has been found to be proportional to the number of people aware of it [56].

In the decision making literature, rather than the forecasting literature, alternative group-based methods for improving decision making have been proposed and tested. We describe these approaches next and then re-formulate these approaches to aid the anticipation of rare events. Schweiger et al. [57] discuss alternative approaches to engender debate and evaluation of decisions in management teams. They differentiate (i) dialectical inquiry and (ii) devil's advocacy. Both methods systematically introduce conflict and debate by using sub-groups who role-play. In dialectical inquiry, the sub-groups develop opposing alternatives and then come together to debate their assumptions and recommendations. In devil's advocacy, one subgroup offers a proposal, while the other plays devil's advocate, critically probing all elements and recommendations in the proposal. Both methods encourage groups to generate alternative courses of action and minimise tendencies towards premature agreement or convergence on a single alternative. Both methods also lead to a more critical evaluation of assumptions by providing mechanisms for encouraging dissent whilst at the same time fostering a high level of understanding of the final group decision. Nevertheless, these role-played, conflict-enhancing, interventions for improving decision making need to be focussed on factual information because personalities can, inappropriately, become the focus of discussion. Advocates of the techniques argue that they are most-suited to ill-structured non-routine decisions. An empirical study [57] compared both techniques to a non-adversarial approach where decisions were simply discussed with the aim of achieving a consensus amongst group members. Questionnaire ratings by group participants found that the two conflict-based approaches were rated higher in terms of producing better recommendations and better questioning of assumptions. Formalizing and legitimizing conflict can thus enhance perceptions of the quality of the outcome of group decision making. However, whilst conflict can improve perceived decision quality, it may weaken the ability of the group to work together in the future if the role-playing is not sensitively managed. Also, as Nemeth et al. [58] document, authentic minority dissent, when correctly managed, is superior to role-playing interventions in stimulating a greater search for information on all sides of an issue. But, generally the authentic dissenter is disliked even when she/he has been shown to stimulate better thought processes. However, other research has shown that the persistent authentic dissenter, while not liked, can be admired and respected [59]. Also, of course implementation of decisions rests on securing the subsequent cooperation of involved parties and so affective personal criticism invoked in the prior critical debate will be dysfunctional [60].

Yaniv [61] demonstrated the power of role-playing in a laboratory-based study of a framing problem. Here, participants were divided into two groups whose members were psychologically primed with either (i) one or the other of two perspectives on the problem — the heterogeneous condition, or (ii) the same perspective on a decision problem — the homogeneous condition. Each of the two groups then convened to discuss the decision problem and come to a group decision. The results were compelling — the homogeneous grouping revealed a stronger framing effect than the heterogeneous grouping. This minimal manipulation produced a discernable impact on subsequent decision quality. Yaniv concluded that Delphi applications could actively create such heterogeneity by assigning roles to panellists as they make their individual forecasts — such as conservative, pessimistic, or optimistic. Additionally, by our analysis, the roles assigned could include those of an agent provocateur — who provides distinctly different forecasts from those of other panellists and includes a critique of other transmitted rationales within his own accompanying rationale for the forecast — before it is transmitted anonymously to other panellists.

In scenario planning, sometimes scenario development involves a scenario team composed of representatives from multiple agencies. Cairns et al. [62] have argued that the process of scenario planning can provide a non-adversarial, common viewpoint to unite, what may be, fragmented groupings. By contrast, in terms of our analysis, the fragmentation should instead be conserved — at least until the point when any action response to the constructed set of scenarios is debated. In more usual scenario development conducted within a single organization, the conventional process results in the initial development of four skeleton scenarios that are then each fleshed-out by one of four sub-groups. On our analysis, once a particular scenario is fully developed it should then be subjected to adversarial critique by one or more of the other sub-groups. Such a process could also be extended to provide adversarial critiques of the more extreme scenarios whose construction we described earlier in this section.

6. Conclusions

In this paper we reviewed methodologies that aim to aid anticipation of rare, high-impact, events. We examined predictability from the perspective of forecasters' ability to obtain well-calibrated probabilities or point forecasts for events and identified six factors that can lead to poor calibration and hence low predictability. We then examined how successful a range of existing methods are in mitigating these factors, including the use of expert judgment, statistical forecasting, Delphi, prediction markets and scenario planning. We demonstrated that all the extant methods, including combinations of methods, contain weaknesses and that anticipation of rare, high-impact, events can only be achieved by judgmental heuristics that, likely, entail bias. We conclude that the only remedies are to either (i) provide protection for the organization against the occurrence of negatively-valenced events whilst allowing the organization to benefit from the occurrence of positively-valenced events — such protection can involve the creation of redundancy, flexibility, and diversity in an organization's operations and resources, or (ii) provide conditions to challenge one's own thinking — and hence improve anticipation. We outlined how use of components of devil's advocacy and dialectical inquiry can be combined with Delphi and scenario planning to enhance anticipation of rare events.

References

[1] S. Makridakis, R.M. Hogarth, A. Gaba, Forecasting and uncertainty in the economic and business world, Int. J. Forecast. 25 (2009) 794–812.
[2] G. Wright, P. Goodwin, Decision making and planning under low levels of predictability: enhancing the scenario method, Int. J. Forecast. 25 (2009) 813–825.
[3] P. Goodwin, G. Wright, Decision Analysis for Management Judgment, Wiley, Chichester, 2004.
[4] N.N. Taleb, The Black Swan: The Impact of the Highly Improbable, Penguin, London, 2008.
[5] D. Orrel, P. McSharry, Systems economics: overcoming the pitfalls of forecasting models via a multidisciplinary approach, Int. J. Forecast. 25 (2009) 734–743.
[6] D.L. Hamilton, T.L. Rose, Illusory correlation and the maintenance of stereotypical beliefs, J. Pers. Soc. Psychol. 39 (1980) 832–845.
[7] A. Tversky, D. Kahneman, Judgment under uncertainty: heuristics and biases, Science 185 (1974) 1124–1131.
[8] D. Kahneman, D. Lovallo, Timid choices and bold forecasts: a cognitive perspective on risk taking, Manag. Sci. 39 (1993) 17–31.
[9] A. Cooper, C. Woo, W. Dunkelberger, Entrepreneurs' perceived chances for success, J. Bus. Venturing 3 (1988) 97–108.
[10] G. Gigerenzer, Why the distinction between single event probabilities and frequencies is important for psychology (and vice versa), in: G. Wright, P. Ayton (Eds.), Subjective Probability, Wiley, Chichester, 1994.
[11] G. Johnson, Strategic Change and the Management Process, Blackwell, Oxford, 1987.
[12] P.S. Barr, J.L. Stimpert, A.S. Huff, Cognitive change action, and organizational renewal, Strateg. Manage. J. 13 (1992) 15–36.
[13] D. Miller, P. Friesen, Momentum and revolution in organizational adaptation, Acad. Manage. J. 23 (1980) 591–614.
[14] R. Fildes, P. Goodwin, Good and bad judgment in forecasting: lessons from four companies, Foresight 8 (2007) 5–10.
[15] Y. Bensalah, Steps in applying extreme value theory to finance: a review, Bank of Canada Working Paper 2000–20, 2000.
[16] I. Kabus, You can bank on uncertainty, Harvard Bus. Rev. (May–June 1976) 95–105.
[17] A. Hoerl, H.K. Fallin, Reliability of subjective evaluation in a high incentive situation, J. Roy. Stat. Soc. 137 (1974) 227–230.
[18] G. Keren, Facing uncertainty in the game of bridge: a calibration study, Organ. Behav. Hum. Dec. Process. 39 (1987) 98–114.
[19] A.H. Murphy, B.G. Brown, A comparative evaluation of objective and subjective weather forecasts in the United States, in: G. Wright (Ed.), Behavioural Decision Making, Plenum, New York, 1985.
[20] S. Oskamp, Overconfidence in case-study judgements, J. Consult. Psychol. 29 (1965) 261–265.
[21] E.H.A. Wallace, What is the Corn Judge's mind? J. Am. Soc. Agron. 15 (1923) 300–304.
[22] P.E. Tetlock, Expert Political Judgment, Princeton University Press, Princeton, 2005.
[23] F. Bolger, G. Wright, Assessing the quality of expert judgment: issues and analysis, Decis. Support Syst. 11 (1994) 1–24.
[24] G. Rowe, G. Wright, Differences in expert and lay judgements of risk: myth or reality? Risk Anal. 21 (2001) 341–356.
[25] R.J. Maxwell, The British government's handling of risk: some reflections on the BSE/CJD crisis, in: P. Bennett, K. Calman (Eds.), Communications and Public Health, Oxford University Press, Oxford, 1999, pp. 94–107.
[26] G. Wright, K. van der Heijden, R. Bradfield, G. Burt, G. Cairns, The psychology of why organizations can be slow to adapt and change: and what can be done about it, J. Gen. Manage. 29 (2004) 21–36.
[27] F. Bolger, G. Wright, G. Rowe, J. Gammack, R.J. Wood, LUST for life: developing expert systems for life assurance underwriting, in: N. Shadbold (Ed.), Research and Development in Expert Systems, VI, Cambridge University Press, Cambridge, 1989.
[28] P. Goodwin, G. Wright, Improving judgmental time series forecasting: a review of the guidance provided by research, Int. J. Forecast. 9 (1993) 147–161.
[29] M.E. Paté-Cornell, Fault trees versus event trees in reliability analysis, Risk Anal. 4 (1984) 177–186.
[30] M.H. Bazerman, M.D. Watkins, Predictable Surprises, Harvard Business Press, Boston, 2008.
[31] I.L. Janis, L. Mann, Victims of Groupthink, Houghton Mifflin, Boston, 1972.
[32] W.Y. Lee, P. Goodwin, R. Fildes, K. Nikolopoulos, M. Lawrence, Providing support for the use of analogies in demand forecasting tasks, Int. J. Forecast. 23 (2007) 377–390.
[33] K.C. Green, J.S. Armstrong, Structured analogies for forecasting, Int. J. Forecast. 23 (2007) 365–376.

[34] R. Fildes, P. Goodwin, Against your better judgment? How organizations can improve their use of management judgment in forecasting, Interfaces 37 (2007) 570–576.

[35] D.S. Turner, The role of judgment in macroeconomic forecasting, J. Forecast. 9 (1990) 315–345.

[36] R. Fildes, P. Goodwin, M. Lawrence, K. Nikolopoulos, Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning, Int. J. Forecast. 25 (2009) 3–23.

[37] M.R. Donihue, Evaluating the role judgment plays in forecast accuracy, J. Forecast. 12 (1993) 81–92.

[38] T.R. Willemain, The effect of graphical adjustment on forecast accuracy, Int. J. Forecast. 7 (1991) 151–154.

[39] P. Goodwin, R. Fildes, M. Lawrence, K. Nikolopoulos, The process of using a forecasting support system, Int. J. Forecast. 23 (2007) 391–404.

[40] D. Önkal, M.S. Gönül, Judgmental adjustment: a challenge to providers and users of forecasts, Foresight 2 (2005) 13–17.

[41] G. Rowe, G. Wright, Expert opinions in forecasting; the role of the Delphi technique, in: J.S. Armstrong (Ed.), Principles of Forecasting: A Handbook for Researchers and Practitioners, Kluwer Academic Publishers, Boston, 2001, pp. 125–144.

[42] G. Rowe, G. Wright, The impact of task characteristics on the performance of structure group forecasting techniques, Int. J. Forecast. 12 (1996) 73–89.

[43] S.J. Grossman, On the efficiency of competitive stock markets where traders have diverse information, J. Finance 31 (1976) 573–585.

[44] J. Wolfers, E. Zitzewitz, Prediction markets in theory and practice, in: S.N. Durlauf, L.E. Blume (Eds.), The New Palgrave Dictionary of Economics, 2nd ed., Palgrave Macmillan, London, 2008.

[45] D.M. Pennock, S. Lawrence, C.L. Giles, F.A. Nielsen, The real power of artificial markets, Science 291 (2001) 987–988.

[46] K. van der Heijden, R. Bradfield, G. Burt, G. Cairns, G. Wright, The Sixth Sense: Accelerating Organisational Learning with Scenarios, Wiley, Chichester, 2002.

[47] R. Bradfield, G. Wright, G. Burt, G. Cairns, K. van der Heijden, The origins and evolution of scenario techniques in long range business planning, Futures 37 (2005) 395–812.

[48] G. Wright, G. Cairns, P. Cairns, Teaching scenario planning: lessons from practice in academic and business, Eur. J. Oper. Res. 194 (2009) 323–335.

[49] R. Phelps, C. Chan, S.C. Kapalis, Does scenario planning affect performance? Two exploratory studies, J. Bus. Res. 51 (2001) 223–232.

[50] S. Makridakis, N. Taleb, Decision making and planning under low levels of predictability, Int. J. Forecast. 25 (2009) 716–733.

[51] N.N. Taleb, Errors, robustness, and the fourth quadrant, Int. J. Forecast. 25 (2009) 744–759.

[52] P. Goodwin, G. Wright, Enhancing strategy evaluation in scenario planning: a role for decision analysis, J. Manag. Stud. 38 (2001) 1–16.

[53] C.M. Pearson, J.A. Clair, Reframing crisis management, Acad. Manag. Rev. 23 (1998) 59–76.

[54] D.M. Schweiger, W.R. Sandberg, J.W. Ragan, Group approaches for improving strategic decision making: a comparative analysis of dialectical inquiry, devil's advocacy, and consensus, Acad. Manage. J. 29 (1986) 51–71.

[55] I.L. Janis, Victims of Groupthink, Houghton Miflin, Boston, 1972.

[56] G. Stasser, D. Stewart, Discovery of hidden profiles by decision making groups: solving a problem versus making a judgment, J. Pers. Soc. Psychol. 63 (1992) 426–434.

[57] D.M. Schweiger, W.R. Sandberg, P.A. Rechner, Experiential effects of dialectical inquiry, devil's advocacy, and consensus approaches to strategic decision making, Acad. Manage. J. 32 (1989) 745–772.

[58] C. Nemeth, K. Brown, J. Rogers, Devil's advocate versus authentic dissent: stimulating quantity and quality, Eur. J. Soc. Psychol. 31 (2001) 707–720.

[59] C. Nemeth, C. Chiles, Modeling courage: the role of dissent in fostering independence, Eur. J. Soc. Psychol. 18 (1998) 275–280.

[60] A.C. Amason, Distinguishing the effects of functional and dysfunctional conflict on strategic decision making: resolving a paradox for top management teams, Acad. Manage. J. 39 (1996) 123–148.

[61] I. Yaniv, Group diversity and decision quality: amplification and attenuation of the framing effect, Unpublished manuscript (2009).

[62] G. Cairns, G. Wright, R. Bradfield, K. van der Heijden, G. Burt, Enhancing foresight between multiple agencies: issues in the use of scenario thinking to overcome fragmentation, Futures 38 (2006) 1011–1025.

Paul Goodwin is a Professor of Management Science at the University of Bath, UK. He is an associate editor of the International Journal of Forecasting.

George Wright is a Professor at Durham Business School, UK. He is editor of the Journal of Behavioral Decision Making and an associate editor of both the International Journal of Forecasting and the Journal of Forecasting.