Can Narrative-based Scenarios Support Quantitative Judgmental Forecasting?

Paul Goodwin [a*]

George Wright [b]

a School of Management, University of Bath, Claverton Down, Bath BA2 7AY, UK

Email: p.goodwin229@btinternet.com

b Strathclyde Business School, University of Strathclyde, 199 Cathedral St, Glasgow G4 0QU, UK.

Email: george.wright@strath.ac.uk

*Corresponding author

RUNNING TITLE: Scenarios and forecasting

KEYWORDS. Judgmental Forecasting; Scenario planning; Narratives; Decomposition; Uncertainty

Can Narrative-based Scenarios Support Quantitative Judgmental Forecasting?

Abstract

Narrative-based scenario planning and forecasting are often regarded as distinct methods for informing decisions subject to risk and uncertainty. This paper compares the approaches and explores the extent to which narrative scenarios can enhance quantitative judgmental forecasts. It argues that scenarios can provide a transparent rationale and context for forecasts, thereby increasing their acceptability. While there is little extant evidence that scenarios can be effective in mitigating judgmental biases in forecasting, this may result from the abbreviated form of the scenarios provided and the non-involvement of forecasters in their development. However, the integration of quantitative forecasting models with scenarios can enhance the former's value - by exposing inconsistencies and discrepancies that may require resolution, and revealing underlying forecast assumptions that need to be both appreciated and monitored.

KEYWORDS. Judgmental Forecasting; Scenario planning; Narratives; Decomposition; Uncertainty

Introduction

Narrative-based scenario planning and forecasting are often regarded as distinct methods for informing decisions that are subject to future risk or uncertainty. In the foresight literature there has been little interchange of ideas between the two approaches with the established forecasting journals, such as the *International Journal of Forecasting* and the *Journal of Forecasting* providing little coverage of scenarios, while journals such as *Futures* have focussed more on scenarios. This may because scenario planning was primarily developed to support longer-term strategic decision-making. In contrast, in forecasting the focus has been on decision-making for the short- to medium-term future. The approaches also differ in their presentation of plausible futures. Scenarios presented as narratives describe how given futures might unfold and they can also explain why. These narratives are multivariate in that they describe plausible outcomes across a range of factors, such as economic growth, demographic changes, and actions of competitors and governments. A key component of the narrative underpinning is the causality of events – how the occurrence of an event earlier in the scenario storylines influences the occurrence of events later in the scenario storyline - though scenarios may also reflect feedback loops[1]. Forecasts, on the other hand, are often quantitative and unidimensional (e.g., first year sales of a new product will be 5000 units or Technology X will take eight years to develop), or they specify a particular outcome (e.g., Technology Y will not be successful).

Despite these differences and others that we identify later, the methods have some commonalities. In particular, the effectiveness of both methods can be highly dependent on

[1] Derbyshire and Wright (2017) provide a nuanced view of causality and feedback loops in underpinning scenario influence diagrams. See also Kunc (2024) for a detailed systems dynamics approach to scenario development.

the quality of human judgment. Moreover, many conditions beneficial for forecasting, such as open-mindedness, inquisitiveness, reflectiveness, and a willingness to debate amongst diverse groups (Tetlock and Gardner, 2016), are also likely to benefit scenario planning. While forecasting can be algorithm-based, there is much evidence that judgmental forecasts predominate in most organizations, either as a substitute for algorithms or as a means of adjusting algorithmic forecasts (Fildes & Goodwin, 2007; Fildes & Petropoulos, 2015). Judgment can be particularly important in forecasts involving new technology and new products where past data on existing technologies may be of limited value and preclude the use of algorithms. Scenarios commonly rely exclusively on judgment and the pivotal role of judgment may continue despite developments in AI. Spaniol and Rowland (2023) recently explored the use of AI in scenario development processes. They concluded that, while AI can potentially have a useful role, it is unlikely to eclipse the need for judgment by scenario planners. Given that the quality of judgment in forecasting can be impaired by widely documented cognitive biases (Karelse, 2021), we investigate whether there is scope for combining the two approaches to mitigate these biases.

We first provide working definitions of forecasts and narrative-based scenarios, before identifying the main differences between the approaches. We then explore the extent to which scenario planning can provide support for forecasting based on a review of studies where researchers have examined the effect of integrating the two approaches. Finally, we identify areas where further research is likely to be beneficial.

Forecasts and Scenarios

Forecasts have been subject to several definitions. The Cambridge Dictionary defines a forecast as 'a statement of what is judged likely to happen in the future'. Clements and Hendry (2002, p.2) adopt a broader definition, indicating that 'a forecast is any statement about the future'. For our purposes, to distinguish between forecasts and scenarios, we offer a more precise definition.

> 'A forecast is an honest estimate of what will happen in a future period where there is a lack of certainty. It is dependent on assumptions made at the time it is created. Formally, it is represented as a probability distribution, but it may be abbreviated to a measure of central tendency such as a mean or mode, a quantile, or a prediction interval.' (based on Goodwin, (2023)).

Based on this definition, a point forecast is merely a central tendency of a probability distribution. Forecasts can also be made for events, which can have probabilities attached, such as the date when a new technology will be available.

Reaching agreement on a definition for the term' scenario' poses a challenge, according to Spaniol and Rowland (2019), who assessed over 400 full and partial definitions. This partially reflects the diverse approaches to scenario thinking that have been practiced. However, Spaniol and Rowland's analysis suggested that broad agreement could be reached on the following attributes of scenarios:

> '...scenarios primarily have a temporal property rooted in the future and reference external forces in that context; scenarios should also be possible and internally plausible while taking the proper form of a story or narrative description; scenarios seem to exist in sets, and the scenarios that inhabit those sets are systematically prepared to co-exist as meaningfully different alternatives to one another.'

In practice, many different forms of descriptions of the future are referred to as scenarios. Some forms are predominately quantitative, and these may be based on equation-based models or simulations (Döll and Vassolo, 2004). Qualitative scenarios, on the other hand are characterised by narrative descriptions of plausible futures. These may simply be short descriptions that outline a particular future without tracing its pathway or explaining its underlying rationale. Other scenarios describe a desired future so that the means of achieving this can be identified (e.g., Kishita et al. 2017). Bunn and Salo (1993) classify different types of scenarios in terms of their intended purpose. They distinguish between (i) scenarios that are designed for management development (e.g. using gaming or role-playing exercises), (ii) scenarios intended to allow the testing of the performance of a proposed strategy against alternative futures, and (iii) scenarios supporting the development of a range of forecasts where a decision must be made between alternative strategies.

It is the narrative form of the third type of scenario that we explore here. A number of methods have been developed to obtain narrative-based scenarios, including the widely applied Intuitive Logics (IL) method (Cairns and Wright, 2018) (see appendix). These methods are characterised by (i) the participation of individuals with diverse expertise and experience and a facilitator, (ii) a structured and iterative process of elicitation, (iii) the identification of key drivers of change whose future trajectories are known, (iv) the identification of key drivers with uncertain futures, and (v) the development of a set of alternative storylines that describe the possible interaction of these drivers over time. These storylines are collectively intended to bound the range of possible futures, taking into account their impact on the development team's focal issue of concern. Each scenario is an internally consistent and plausible narrative providing a multidimensional account of how and why a future might unfold from the current point in time to a horizon year. Importantly, it is the multidimensional space between the scenarios (of which there are usually four developed when applying the IL method) that serves to represent and highlight the degree of uncertainty faced. The greater this space, the greater the uncertainty. Note that the IL method and others do not usually take the desirability of a to-be-developed scenario into account, i.e., they do not aim to obtain best- and worst-case scenarios within, what usually is, an "open-ended" scenario development process (see Appendix for detail of a typical IL process).

Contrasting forecasting and narrative scenario planning.

Some researchers have argued that there has been a convergence of scenario planning and forecasting. For example, writing in 1993, Bunn and Salo (1993) concluded that 'its [scenario analysis's] time as a distinct alternative to forecasting and planning may have already passed.' Others regard the two approaches as serving very different purposes. Van der Heijden (2005) argued that 'Scenarios are not seen as quasi-forecasts but as perception devices' providing a 'means of thinking'. He also regarded forecasting as applicable only to the short-term future where predictability is high, while scenarios applied to the medium term where there is a level of predictability but also considerable uncertainty.

Table 1 summarises the main differences between forecasting and scenario planning, as defined earlier and as commonly practiced. Detailed accounts of the practice of forecasting in companies can be found in case-study papers, such as Fildes and Goodwin (2021), while survey-based papers, such as Fildes and Goodwin (2007) and Goodwin et al. (2023) provide evidence of practices which are common across organizations. Fildes and Stekler (2002) show that many of these practices are also evident in macroeconomic forecasting.

While a scenario will tell a story of the unfolding events on a pathway to a given future, forecasts will only present a future as a finally resolved outcome, be it an event or a number. While the future presented by a scenario will be multidimensional - i.e., where the intersection of unfolding events (e.g., including political, economic, social, technological, environmental, and legal/regulatory events) reaches its culmination in the scenario 'horizon year', a forecast will usually be unidimensional. Crucially, a scenario provides a documented causal rationale for the plausibility of a given future. By contrast, in forecasting, the underpinning rationale is often not made explicit or recorded, or proffered rationales may be spontaneous and unstructured (Fildes and Goodwin, 2021). However, in judgmental forecasting methods such as Delphi, good practice involves panellists being asked to provide a sentence or two of justifications for their judgmental forecasts. (Rowe and Wright, 1999). Explanatory statistical forecasting methods such as multiple regression can also provide such a rationale, albeit implicitly. However, the story-telling approach of scenarios is a much more natural, and to many people a more amenable, way of understanding the world and how it may develop than a mathematical model. For example, Taleb (2007) has stated that "Statistics are invisible; anecdotes are salient". Beach (2021) has argued "… the brain organizes events into a temporal/causal narrative of the past and present and extrapolates a plausible version of the future as a matter of course—automatically, if you will." Similarly, Bulley et al., (2022) argue, from a review of the developmental studies that have been conducted by psychologists, that many of the key competencies and capabilities for futures thinking (rather than forecasting) are present in adults, suggesting that that humans think causally and are natural scenario planners, rather than forecasters.

Insert Table 1 about here

While scenarios attempt to describe plausible (i.e., causally based) extremes that embrace a range of plausible futures, in practice, forecasts are often confined to the most likely (modal) outcome or the mean outcome of the underlying probability distribution. Such forecasts provide no guidance on the risk or uncertainty associated with a decision. However, when probabilities do accompany them, forecasts can be useful in decisions such as determining where resources should be allocated now, in order to prepare for, what are seen as, more, or less, likely future events (e.g., a predicted degree of upturn in the sales of a particular promoted product or service). See Goodwin and Wright (2014) for details of probability-based decision analysis methods and approaches to aiding decision making under uncertainty. Nevertheless, estimating the required probabilities without historical data can be problematic, while judgmental probabilities can be subject to biases arising from heuristics such as availability and representativeness (Tversky and Kahneman, 1974).

By contrast, scenarios generated by methods such as the IL process are not, in practice, associated with probabilities or measures of their relative plausibility. All outcomes in the multidimensional space between the set of scenario 'extremes' are usually treated as being equally plausible. This means that no specific guidance is provided on where resources should be preferentially directed – in order to deal with a potential future that may be negatively or positively valanced (Scoblic & Tetlock, 2020) - to make a sound decision now, in the face of future uncertainty. As a result, while forecasting is focused on supporting the identification of optimal decision options, scenarios aim to help decision-makers design robust options in the face of the alternative futures captured "between" the set of four scenarios. See Goodwin and Wright (2001), Durbach (2019), and Goodwin (2019) for more details on the combination of scenario thinking with non-probability-based decision analysis methods.

As we have discussed, in practice, both forecasting and scenario development are heavily dependent on judgment. However, while in scenario development, the underpinning judgment is elicited by a process based on a well-defined structured decomposition, the use of both structured elicitation and decomposition of judgment is much rarer in forecasting practice. Here, potential structures for the elicitation and structured decomposition/re-composition are both (i) less clearly defined and (ii) little utilized. Where decomposition is applied in forecasting, it most commonly involves delegating part of the forecasting task to an algorithm with informal judgmental adjustments made for information not accounted for. However, such informal adjustments are often applied inefficiently and may involve double counting or the inappropriate use of non-diagnostic cues (Fildes et al., 2019; Fahimnia et al., 2023). In prescriptive decomposition methods, like decision analysis, judgmental probability forecasts are estimated separately from consequences (or utilities), so, unlike scenario methods such as IL, judgments about the future are made without explicit consideration of the impact of outcomes, which should simplify the judgmental task.

Importantly, algorithm-based forecasts rely exclusively on history since they assume constancy - that is, that patterns identified in past data will continue into the future. This may also be true, to some extent, of the focus of the cognitive processes underpinning the development of pure judgmental forecasts, though there is evidence that these may also be subject to a recency bias, with an over-emphasis on the most recent observations (Lawrence & O'Connor, 1992; Bolger & Harvey, 1993) or recent events (e.g. Niu, 2023), and a failure to distinguish between persistent patterns and noise. In contrast, the focus of scenario planning has been to downplay history, relegating it to a peripheral role (Bradfield et al., 2016) so that the emphasis is on identifying the pathway of events from the present to the future. But, even here, there may be a recency bias in that undue attention may be paid to current (e.g., media salient) issues when scenarios are being constructed (O'Brien, 2004; Wright et al., 2009).

Scenario development is often a team-based workshop exercise over two to three days, or even much longer (Bennett & Zurek, 2006). The team members are chosen because they bring different perspectives (and thus may identify different driving forces) that are discussed and integrated during the workshop (see Cairns and Wright, 2018, chapter 2). Because of the need for team-based activity and the intricacies of the staged development process, scenario development is both expensive and time-consuming and, hence, is likely to be carried out relatively infrequently. Given its broad-brush, multi-attribute presentation of plausible futures as formulated scenarios that will unfold from the present point-in-time to the scenario horizon year, it will also be difficult to establish that the future, when it is finally resolved as a set of outcomes, falls neatly within the span of futures that were developed earlier as anticipatory scenarios. This is because the actual broad-brush future will necessarily include unexpected events. Because of the long-term focus of scenario storylines, such an exercise would probably have little value anyway, as it could only be made years after the scenarios were first constructed and associated decisions made. Indeed, the key benefits of many scenario planning exercises may emanate from the scenario construction process itself, in that it serves to sensitize decision-makers to the uncertainties that they face and the prospects for fundamental change in the contextual environment so that strategic inertia is avoided (Goodwin and Wright, 2014).

By contrast, in forecasting, there is an established set of bias and accuracy metrics, but all have their limitations (Makridakis, 1993; but also see Koutsandreas et al., 2022). Their reliability is limited by the number of observations in the out-of-sample data (i.e., the observations that were not known at the time the forecast was made, and which are used to test forecast accuracy) (Goodwin, 2011) and the assumption that performance in relation to these observations will be reflected in the future. However, as in scenario planning, judgmental forecasting processes, such as sales and operations planning (S&OP), may, if well conducted, yield benefits in themselves in that they allow for a pooling of information and opinions from different sources and the potential to challenge the views of others.

Finally, algorithm-based forecasts, such as linear and logistic regression, have clear theoretical underpinnings, while a number of psychological theories underpin judgmental forecasts and improvement strategies. In contrast, scenario planning has been criticized for lacking a theoretical base. However, Derbyshire (2017) has argued that Shackle's Potential Surprise Theory can provide theoretical support for the IL scenario planning practice.

Scenarios as support for forecasting

Arguments favouring the use of scenarios to support forecasts

Are the differences between forecasting and scenario planning irreconcilable? Or can the two approaches complement each other so that decision-makers can draw on deeper insights and improved guidance on what may prevail in the future? We first consider the arguments that favour the use of scenarios as part of the forecasting process, then we consider why the extant evidence suggests that many of the hypothesised advantages of using scenarios do not apply.

The first argument is that scenarios provide a rationale for forecasts which is often absent or not made explicit at the time that a forecast is made. A lack of a linked rationale can reduce the acceptability of a forecast to decision makers (Taylor and Thomas, 1982; Gregor and Duran, 2001). Moreover, the process of creating a rationale for a focal forecast as a single "business-as-usual" contextual scenario may be likely to improve the accuracy of the forecast because it will force forecasters to confront their implicit assumptions at the time that a forecast is made, assumptions which may be invalid (Karelse, 2021). When assumptions are unchallenged, they can result in forecasts being based on unsubstantiated rumours or non-diagnostic cues (Fildes et al., 2019), or they can result in gratuitous judgmental adjustments of algorithm-based forecasts (Goodwin, 2000). Informal assumptions made at the time a forecast is produced may also be difficult to recall at some future point in time when subsequent forecast accuracy may be being evaluated, thereby restricting the scope for learning. In contrast, by providing an explicit rationale for why a given contextual, broad-brush, future will prevail, the assumptions underlying the focal forecast are made transparent and open to both contemporaneous challenge and post-hoc evaluation.

Judgmental forecasts are subject to a number of well documented cognitive biases several of which are interrelated. For example, recent observations and events are often over weighted when estimating what the future will hold (Bolger and Harvey, 1993). This can result from the use of the availability heuristic, where easily recalled events are judged to be more probable in the future, or the anchor-and-adjust heuristic where change from the most recent observation, which acts as an anchor, tends to be underestimated. There is also the end-of-history illusion which perceives that the current state is stable, and the future will be a linear continuation of this state (Schirrmeister et al., 2020). In contrast, people sometimes perceive systematic changes in past data which merely reflect noise. The representativeness heuristic (Tversky and Kahneman, 1974) causes them to assume that such perceived pattens are representative of future conditions. As a result, they may make unjustified adjustments to statistical forecasts that project stable continuity in what is, in fact, noisy data. Optimism bias (Fildes et al., 2009) is also common in forecasting. While this may reflect motivational biases, such as organizational politics, it can also result from wishful thinking (Goodwin and Fildes, 2022).

It can be argued that by focusing attention on how driving forces might play out and interact in the future, the utilisation of multiple contrasting scenarios will reduce the focus on recent events and a singular business-as-usual view of the world (Schoemaker, 1991). In addition, the presence of worst-case scenarios could dampen the tendency for wishful thinking and other causes of unjustified optimism (Gregory and Duran, 2001, p523). Similarly, the rationale provided by scenarios might be expected to reduce the tendency to project, what may be, random fluctuations seen in past observations into the future. Many biases are believed to result from the limited processing capacity of the human mind (Hogarth, 1987) causing people resort to heuristics, or simplifications, when making judgments. The decomposition of the judgmental task inherent in the scenario elicitation process, it can be argued, will reduce the cognitive burden on forecasters and hence reduce the reliance on heuristics (MacGregor, 2001).

People also tend to underestimate uncertainty, resulting in overconfidence (Ferretti, Montibeller & von Winterfeldt, 2023). As a result, forecasts presented as prediction intervals are often too narrow given their stated probability of capturing the actual outcome (Glaser et al, 2013). Given that scenarios come in multiples, typically of four, with each based on alternative assumptions, they can provide a sense of the span of plausible futures and hence aim to sensitize decision-makers to the degree of uncertainty they face (Schoemaker, 1991).

The extent to which uncertainty can be recognized for forecasts based on contextual scenarios is reflected in Figure 1. Figure 1(a) consists of a point forecast of central tendency and estimates of bounds based on worst- and best-case scenarios. Note that the bounds depicted in Figure 1(a) are essentially point forecasts conditional on each of two "extreme" scenarios. However, these point forecasts of extremities may themselves be subject to uncertainty, and Figure 1(b) shows a more complex representation with bounds on each of *these* point forecasts. In the following discussion, we will refer to these two representations as first and second-order uncertainty, respectively.

Insert Figure 1 about here

In addition to biases of individual forecasters, group dynamics can adversely affect forecast accuracy. For example, many important forecasts in companies are made by managers attending forecast review meetings or engaging in sales and operations planning (S&OP) processes (Fildes and Goodwin, 2021). Status differentials, pressures to conform and phenomena such as groupthink can diminish the forecast accuracy of any forecast agreed by a group and exacerbate the tendency to underestimate uncertainty. Carefully structured scenario processes with well-trained facilitators who are aware of these dangers might be expected to mitigate these social context problems (Schirrmeister et al., 2020).

In summary, the idea of integrating scenarios and forecasting has many parallels with mixed methods research and therefore could potentially yield similar benefits. These include: (1) balancing out the shortcoming of each method by combining different approaches; (2) providing stronger evidence and more confidence in findings through iteration; (3) combining numeric with narrative approaches to avoid over-reliance on the former and enabling the capture of "soft-core views and experiences" which elucidate complex social situations (Jogulu and Pansiri, 2011). However, the ontological and epistemological assumptions of both judgmental forecasts and scenarios are arguably similar – they assume a relativist ontology and a subjectivist epistemology. This may restrict the extent to which hoped-for benefits are obtained, and we next examine extant studies suggesting that the quality of forecasts is not improved by attempting to support their production with scenario thinking.

Counter evidence

Despite the above arguments, Gregory and Duran (2001) concluded that ‘no evidence exists that scenarios can be used to develop more accurate forecast for economic events.’ This reflected the findings of a laboratory study conducted by Schnaars and Topol (1987) which found that judgmental forecasters who had access to multiple scenarios produced did not produce more accurate point forecasts than those not provided with scenarios. In both cases, the judgmental forecasts were less accurate than those provided by statistical methods. The effect of the scenarios was to make the forecasters more confident about their predictions - which appeared to be based on a single favoured scenario rather than the entire set. However, the latter finding is perhaps not surprising. The participants who received scenarios in the experiment were provided with a time-series graph showing the annual sales of a product over 11 years and three scenarios described as optimistic, pessimistic and middle ground. It is unclear how they were expected to integrate the information in these scenarios to estimate a point forecast, but an obvious strategy would be to examine the recent behaviour of the series and identify which scenario was most closely aligned to that movement. Indeed, this is what happened in a study by Goodwin et al. (2019a). In this study, forecasters were also shown a time series of product sales alongside best-case and worst-case scenarios (a first-order representation), which suggested market conditions that might cause future sales to increase or decrease. Rather than mitigating the bias where too much attention is paid to the most recent change in the series (Bolger and Harvey, 1993), having access to scenarios exacerbated this bias. If there was a recent rise in the series, a best-case scenario focused attention on the sales increase, while the worst-case scenario was discounted. Similarly, the presence of a worst-case scenario led to greater attention being paid to a recent fall in sales, with the best-case scenario being disregarded.

While the above findings suggested that access to scenarios can be damaging for point forecasts made when time series data are available, it was suggested above that access to extreme scenarios can be effective in reducing the tendency of forecasters to underestimate uncertainty (i.e. be subject to overconfidence). This would be the case if worst- and best-case scenarios alert forecasters to the need to estimate more extreme end points in prediction intervals. However, estimating each end point would require the forecaster to focus on each scenario separately so their judgment is not contaminated by the presence of the alternative scenario. Goodwin et al. (2019b) conducted a laboratory experiment where participants in a control group were provided with *either* a best or worst-case scenario and asked to make point and interval forecasts of demand for a product, assuming that the provided scenario would prevail (a second-order representation). Other participants were provided with *both* best- and worst-case scenarios but asked to make forecasts under the assumption that only one of these scenarios would prevail. For example, they were asked to make forecasts given that a best-case scenario would happen, but they were also provided with a worst-case scenario, which should have been irrelevant to their judgment. However, the effect of having access to the opposite scenario led to a narrowing of prediction intervals (i.e., greater overconfidence) for forecasts based on the relevant scenario, the opposite of what was desirable, with the effect being greater where scenarios were more extreme. This result is consistent with earlier studies finding that confidence in forecasts increases when more *information* is available, even when the extra information is irrelevant (e.g., Hall et al., 2007; Oskamp, 1965).

The above finding, that prediction intervals based on a best-case scenario are influenced by the provision of an irrelevant worst-case scenario, raises the possibility that the presence of a worst-case scenario might temper any optimism bias found in *point forecasts* made for a best-case scenario (this would also suggest that a point forecast assuming a worst-case scenario might be made more optimistic when an irrelevant best-case scenario is available). Despite this, the study by Goodwin et al. (2019b) (using a second-order representation) found that the opposite was the case - point forecasts made on the assumption that a given scenario would prevail appeared to be subject to a contrast effect (Sherif et al., 1958). When an irrelevant opposing scenario was also provided, point forecasts based on the original scenario became more extreme. This meant that forecasts based on a best-case scenario became more optimistic when a worst-case scenario was also provided. However, forecasts based on a worst-case scenario became more pessimistic when an irrelevant best-case scenario was available. Again, the effect was greater where more extreme scenarios were available. This is consistent with the psychology literature suggesting that contrast effects are more likely to occur where differences between the target stimulus (in this case, the scenario assumed to prevail) and the context (the opposing scenario) are greater (Sherif et al., 1958). Where the two stimuli are less distinct and may even be perceived as overlapping an assimilation effect is more likely. In the case of scenarios, this would imply that a point forecast based on a 'moderate' best-case scenario would be less optimistic if a worst-case scenario were also provided and vice versa. (Chien, et al., 2010).

Forecasts obtained from computer-based algorithms are often subject to an excessive number of judgmental interventions, leading to a reduction in accuracy (e.g. Fildes et al., 2009; Fildes & Goodwin, 2021; Franses and Legestee, 2010). In practice, forecasters in some companies adjust over 90 per cent of their computer-based demand forecasts. Many of these adjustments are small and may be made to justify the forecaster's role or to give them a sense of ownership of the forecasts (Önkal and Gonul, 2005). As such, they cause little harm to accuracy but can waste the forecaster's time. In cases where the forecaster has information that is not available to the algorithm and is likely to have a significant impact on the variable being forecast, such adjustments are likely to be beneficial, but occasions like this are likely to be relatively rare.

Can the provision of scenarios reduce this tendency for excessive adjustment? This question was examined in a study by Önkal et al. (2013), which was based on a first-order representation of uncertainty. Participants were provided with time series data and algorithm-based one-period-ahead point forecasts of the demands for a series of products. Depending on the treatment they were randomly assigned to, they received either no scenario, a worst-case scenario, or both worst and best-case scenarios. After they used the provided information to make their own judgmental point and best and worst-case forecasts, the participants met in groups of two to agree on consensus forecasts. The provision of opposing scenarios led to the judgmental point forecasts of central tendency being closer to the model-based point forecasts both in the case of individuals and groups -an outcome that is likely to be associated with greater accuracy. However, the benefits of scenario provision were not unmitigated. Providing either single or opposing scenarios generally led to implicitly narrower prediction intervals because people also stayed closer to the algorithm-based point forecasts when assessing the bounds of possible demand. Hence, access to scenarios led to an increase in overconfidence.

In summary, the limited evidence available suggests that narrative-based scenarios are ineffective, both in improving the accuracy of judgmental point forecasts and in reducing the overconfidence associated with judgmental prediction intervals. In particular, when provided with a set of scenarios, forecasters tend to focus on a single scenario that appears to be favoured by recent observations. Conversely, when asked to assume that a particular scenario will prevail, their forecasts are influenced by other scenarios in the set, which are irrelevant. However, in the next section we evaluate the extent to which these conclusions are supported by this extant laboratory-based evidence.

Examining the counter evidence

It is important to note that many of the above studies had the valid aim of examining how diverse and opposing pieces of qualitative information and opinions interact and impinge on the judgmental forecasting process. Contrasting information such as this is often available to forecasters. For example, it may be tabled and discussed in forecast review meetings where executive meet to agree important forecasts, though it may be presented verbally and in an unstructured informal way (Fildes & Goodwin, 2021). However, do the above studies provide sufficient evidence to conclude that scenarios damage both accuracy of point forecasts and the calibration of prediction intervals?

We can categorize narrative scenarios on two dimensions. The first type of scenarios are those that are developed from the judgments of the decision-maker and so are representations of the decision-maker's view of the future. We will term this type of scenario *"self-developed"* (SD), even though the development process is likely to have been supported by the help of a skilled facilitator (see Cairns and Wright, 2018, chapter 10). The second type of scenarios are those developed externally from, and then delivered to, the decision maker and so may not capture the decision maker's own anticipations of the future. We will term this type of scenario *"developed by others"* (DO).

Next, we will refer to two types of scenario content. The first type is scenarios that incorporate driving forces from across the PESTEL dimensions - i.e., political, economic, social, technological, environmental, and legal/regulatory. (c.f., Stage 2 of the basic IL process in the Appendix), which we will term *"full scenarios"* (FS). The second type of scenarios' contents are those that are abbreviated and simplified. We will term this type of scenario *"abbreviated scenarios"* (AS). Scenarios developed by processes such as IL can easily be categorised as both "self-developed" and "full scenarios", i.e., can be defined as SD/FS.

The scenarios provided in the laboratory experiments referred to in the above sub-section tended to be short, i.e., not based on the PESTEL framework, and they were 'received' by the participants - that is, the participant forecasters did not develop them. Hence, the scenarios discussed in the sub-section above can be categorized as AS/DO. Examples of "abbreviated" scenarios that were "developed by others" in extant experiments are given in Table 2. In many cases they merely make positive or negative statements about the future, so the presence of a causal chain is less evident. In particular, it is unclear how a forecaster can utilise contextual expressions in the AS/DO scenarios - such as '[the product having] a combination of necessary features to compete successfully', 'the product's popularity is very high', 'receives positive comments in industry magazines' or 'the future looks bright' - into informed quantitative estimates of the future sales of a product. Informed estimates would require additional contextual information within the proffered scenario, such as the size of the market, the estimated impacts of competitors' actions, or the production capacity of the product's manufacturer. For example, assuming an abbreviated best-case scenario containing minimal contextual statements, all a demand forecaster can do is to simply place his/her estimate somewhere higher than those made for no-change or worst-case scenarios, but it is unclear where. (For greater detail on the quality yardsticks for both scenario development and content, see Crawford and Wright, 2022).

Nevertheless, it should be noted that, despite their possible advantages in providing a detailing forecasting context, full scenarios would need to be used with care. In particular, their use might overload forecasters with information, so they only draw on a part it. This suggests that some form of decomposition would be required (Sjöberg, 1982). Full scenarios might also encourage forecasters to attach excessive credence to the futures they describe, given their compelling detailed underlying rationales (Tversky & Kahneman, 1983; Schoemaker, 1993). The extant AS/DO laboratory studies do not speak to these research issues.

Insert Table 2 about here

Turning to the second dimension, it seems likely that self-developed (SD) scenarios would lead a forecaster to pay greater attention to each scenario's content, including specific information, such as market size, that would be required to produce a forecast. It is also more likely to lead to a more vivid perception of plausible futures (Gregory and Duran, 2001. Beach (2021) has argued that scenarios developed by others may not be assimilated into the decision-maker's own view of the world. He argues that:

> *'the brain organizes events into a temporal/causal narrative of the past and present and extrapolates a plausible version of the future as a matter of course—automatically, if you will. The chronicle of experience this creates is called the prime*

narrative.…Your prime narrative is all you have to understand what is going on; it is your intuitive truth. As a result, it takes a good deal of counter-evidence to make you doubt its truth.'

The issue of the congruence of those scenarios being developed within an ongoing scenario workshop with the participants' own prime narrative is, in our view, an acute one that the facilitator of scenario development within a scenario team must both consider and be able to deal with. In parallel, when scenarios are fully developed by others and presented to a focal decision-maker for evaluation and subsequent action, a challenge to the reader's prime narrative may result in the scenarios being quietly "shelved" or "filed". By contrast, the use of DO/AS scenarios in the laboratory studies that we reviewed earlier would, in our analysis, not be viewed, in any way, as challenges to the reader's (i.e., decision maker's) prime narrative.

Can field applications where scenarios have been used to support forecasts provide evidence of the value of scenarios in forecasting?

Recently, the UK National Grid, which distributes electricity throughout the country, has produced scenarios that 'outline four different pathways for the future of the whole energy system out to 2050' (National Grid, 2023). For example, one scenario, 'Consumer Transformation', envisages a country where the typical household 'will use an electric heat pump with a low-temperature heating system and an Electric Vehicle (EV). The system will have 'higher peak electricity demands managed with flexible technologies including energy storage.' Here, the scenario depicts green-issue-sensitive citizens with money in their pockets. Model-based point forecasts of variables such as greenhouse gas emissions (GHGs) and total electricity and gas demand are then made up to 2050 under the conditions described by each scenario. However, the details of the scenario development process that underpinned National

Grid's four scenarios are not well-specified in their report. There is no reporting of causal scenario storylines in the report, so we categorize the scenario content as abbreviated, i.e. (AS) and little detail is given on the scenario development process (Crawford and Wright, 2022). Further, no details are given about when the forecasts of variables such as electric car sales and heat pump installations were elicited or whether the forecasters were aware of all scenarios at the time of forecasting. From a researcher's perspective, having answers to these and other questions would be helpful. Was a single time series forecast judgmentally adjusted to incorporate the influence of each scenario in turn? How were such adjustments achieved? Were rationales collected? Were the forecasters the same people as those who had previously developed the scenarios? In what order were the scenario-based forecasts elicited?

In contrast to the National Grid project, the Millennium Ecosystem Assessment (2003) did provide details of the process of integrating of scenarios with quantitative forecasts. A part of this project was ‘a scenarios exercise to describe plausible changes in ecosystem services and their consequences for human well-being at a global scale’ (Bennett and Zurek (2006). Each of the four scenarios utilised made different assumptions about how social-ecological systems might develop between 2000 and 2050. The scenario development process was a long one: around seventy experts, drawn from a wide range of academic disciplines, participated in a series of eight workshops over three years - so the scenarios could be categorised as SD/FS. A “storyline-and-simulation” approach was used to achieve integration between the narratives and quantitative forecasts, (Alcamo 2001). This involved translating the causal storylines into numerical variables that served as model inputs. The models were then used to produce forecasts of variables such as water quality, fish harvests and crop production. The process of aligning the model forecasts with the narratives was iterative with comparisons between the two approaches and adjustments to ensure that the results were consistent.

The process of integration adopted was not without its challenges (Bennett and Zurek (2006). The quantification process required the scenarios to be simplified, which led to concerns being raised by participants. Discussions about which aspects of the narratives could be simplified and which needed to remain intact and allowing the outcome of the project to be both presented both as narratives and numbers overcame these problems. The participants also felt that the models alone were not adequate to address the project's focal questions so they were mainly used as a check on consistency and as a framework for the narratives. In particular, while the quantitative models provided scientific credibility, unlike the narratives they were unable to handle nonlinearities. Nevertheless, participants reported that the integration of scenarios and quantitative models allowed the project to incorporate the diverse views of people from quantitative and qualitative disciplines resulting in a richer set of views of the future. Moreover, discussions between them required the defence of assumptions that strengthened the clarity of the project's outcomes.

While the Millennium Ecosystem Assessment project provides useful pointers on how scenarios can provide a context for practice-based forecasting, it was not primarily a forecasting project. Also, because the forecasts were produced using formal models, the project does not provide guidance on how scenarios can support purely judgmental forecasts. The project is also an isolated example of a documented process. The combination of scenarios as contexts for forecasts is relatively recent and, as yet, both under-explored and under-documented.

The practicality of using scenarios to support forecasting of continuous variables.

A number of issues may restrict the practicality of using scenarios to support forecasts. A potential limitation occurs when forecasters want to use the concept of worst- and best-case scenarios. In methods such as IL, scenarios are selected to represent the greatest span of uncertainty and impact. As such, there is no explicit notion of ranking scenarios in terms of likelihood or desirability. It may, of course, be possible to identify, out of the four scenarios that are usually formulated, the two scenarios that offer the best and worst contextual prospects for sales or whatever other variable is being forecast (see our discussion of the use made of the National Grid energy scenarios, earlier). However, given the multiattribute nature of scenarios, this is not a given, and it may be that none of the scenarios developed by the "open-ended" IL process describe situations that provide the absolute worst or best conditions for the forecast variable. As such, approaches like the Backwards Logic Method (BLM) of scenario development (Wright & Goodwin, 2009) may be better attuned to the needs of forecasters seeking to establish a range of possible outcomes for the variable to be forecast. BLM is an adaptation and augmentation of IL designed to broaden the range of scenarios considered while, at the same time, retaining the essential focus on causality within the process of scenario construction.[2] Within the BLM, the focus is shifted onto understanding the causes of plausible changes to the organization's achievement of its objectives, i.e., often, plausible absolute worst or best contextual conditions.

As indicated earlier, where scenarios are confined to extremes that are regarded as equally plausible, their potential value to forecasters is restricted to providing guidance in establishing bounds on the range of possible outcomes rather than support for establishing point forecasts of central tendency.

A third practical issue relates to the resources needed to construct scenarios. Short to medium-term forecasts are produced frequently. Formulating scenarios using methods such as IL or BLM is an expensive and labour-intensive, group-based process, and it would not be

[2] BLM is distinguished from backcasting, as described by Robinson (2003), in that the latter focuses on desirable futures, rather than plausible futures.

practical to produce scenarios on a weekly, monthly, or even a quarterly basis. Hence, such scenarios are likely to be most useful in supporting nationally important longer-term forecasts, such as those produced by the UK National Grid, or significant, influential forecasts made about aspects of a nation's economy. Alternatively, if scenario thinking is normally included as a key component of the strategizing process within an organization, the time and effort devoted to futures thinking may be acceptable. See Rowland and Spaniol, (2022), for a discussion of these issues.

Fourth, there is the issue of how a narrative description of the future can be turned into a quantitative forecast. Several approaches are possible here, including the "storyline-and-simulation" approach that we referred to earlier. A given event described in a FS scenario may, in the past, be associated with a particular level or a specific range of levels in the variable to be forecasted. In this case, forecasts may be based on analogies with earlier outcomes (e.g. Green & Armstrong, 2007) or models that include these events as independent variables. For example, if a given tax increase occurs, it may be estimated to lead to a 12% drop in future demand for a product based on past experience. Market research models and intentions surveys may also be helpful in estimating how people might respond to different events (Morwitz & Munz, 2021). Estimates based on a combination of events in a scenario may be obtained through decomposition. For example, consider a pharmaceutical company estimating a drug's future demand based on a given scenario. A pre-determined trend suggests that the population aged over 65 will increase by 10% by the horizon year. Of these, an estimated 8% would likely have a medical condition and 20% would be prescribed the drug assuming no more advanced treatments are developed, and the regulatory regime remains unchanged. Competing pharmaceutical companies would likely take 30% of the market because the drug will have lost its patent protection by the horizon year. Such procedures have the added advantage of providing a documented rationale for the forecasts that is both transparent and open to informed challenge and debate.

One idea that has been underexplored is the use of scenarios to assess forecasts through 'wind tunnelling'. For example, the plausibility of a forecast made today can be tested against individual scenario storylines. If it is inconsistent with, say, three of the four scenarios then it may need to be reviewed. When forecasts are made in practice, the ambient context is often not systematically explicated or recorded. A structured approach to eliciting a forecast against each of a set of explicated scenarios will: (i) allow the wind tunnelling but will also (ii) provide an audit trail. If an important forecast is a long-term one then, by contemporaneously recording and monitoring the reality of the actual unfolding future, there may be an early warning that the focal important forecast needs to be re-done, since the ambient context is noted to be changing from that contained in the contextual scenario that was utilised at the time of the initial important forecast.

Scenarios as support for judgmental probabilistic forecasting of events

Can scenarios improve probabilistic forecasts of events such as business failings, economic downturns, the winners of elections or future world conflicts? Tetlock (2017) has found that experts in many of these fields rarely produce well-calibrated probability forecasts. It seems possible that engaging them in the construction of scenarios will lead to more profound thought about the likelihood of focal events, allowing them to reflect on the rationale underlying their forecasts and alerting them to the prospects of alternative events occurring. However, there are several obstacles to using scenarios to achieve more accurate judgmental probabilities.

Because the rationales and details provided by scenarios can be persuasive, the likelihood of a specific scenario prevailing can be judged to be unrealistically high, a phenomenon known as simulation bias. Hence, the likelihood of a specific event, conditional on this scenario, may also be overestimated. Moreover, where scenarios represent extremes, they will not be aligned with more moderate conditions that may be more conducive to the occurrence of a given focal event.

Chang et al. (2016) tested an approach they referred to as scenario training to examine how much this improved the calibration of probability forecasts. They hypothesized this would work by 'encouraging trainees to think critically about assumptions, potential futures, and causal mechanisms that could be at play on a given forecasting question.' The training, "never designed to take more than one hour", consisted of "probabilistic reasoning" training and "scenario" training. From the write-up of the study, the elements of the training focussed on: "probability sum rule", "exploring and challenging assumptions", "identifying key causal drivers", and "identification of worst- and best-case scenarios". In short, this training, which was not based on intuitive logics, did improve the accuracy of probability of forecasts of geopolitical events, as measured by Brier scores, when compared to a control group who received no training. Training improved their calibration by reducing overconfidence. However, since the training was short and varied – containing aspects of both probabilistic thinking and scenario thinking – it is unclear which parts were beneficial to improving judgmental forecasting accuracy. Clearly, though, here, the scenarios would have been self-developed (i.e., could be categorized as SD), but whether the content was full (FS) or abbreviated (AS) is unclear. Given the time constraints, our view is that the scenario development activity and scenario content can likely be categorized as SD/AS. More research is needed to establish whether IL-based scenarios that focus on extremes and, hence, highly unlikely future developments can help inform judgmental probability forecasts of events with a sizeable likelihood of occurring.

Discussion and Conclusions

As we have discussed, when judgmental forecasts are made, reasoning underpins the forecast, but the content and extent of this reasoning has been little investigated and is often left implicit and not explicated. For example, in the Delphi process, a focus has been on the exchange of short rationales, but the study of the content of the rationales has largely been left to one side. By contrast, when developing scenarios using methods like IL, the explication and exchange of causal reasoning have been a focus of the structured decomposition of judgment inherent in the process. The IL process is, perhaps, the most used in scenario practice. Here, the scenarios developed for strategic planning are detailed (often 500 words or more) – and the scenarios themselves are often developed by the decision makers themselves (with the help of a seasoned scenario development facilitator). As such, they provide a comprehensive evaluation of plausible changes in the contextual environment across the political, economic, social, technological, environmental and legal/regulatory dimensions. In short, best practice scenario development is both SD and FS.

In parallel, within the experimental laboratory, the impact of scenarios on forecasting has been studied using scenarios that were not developed by the study participants, and that can be characterized as abbreviated – i.e., DO/AS. This situation can have been anticipated since the fields of scenario practice and forecasting research have, until recently, been advancing separately. In short, the situation most amenable to the future integration of the two fields will be when scenario and forecasting researchers work together. Even in recent practice, as exemplified by the National Grid effort, the use of scenarios has been scantily documented and, in our view, rudimentary but is indicative of the scope for future developments and the possibilities for integration.

Since development processes such as those employed in IL are detailed and extensive and explicitly involve the decision-makers themselves, the time and effort needed is considerable. This fact alone indicates that scenarios and forecasts will be best integrated when the forecasts to be made are important – such as nationally relevant long-term economic forecasts. Long-term forecasts, in particular, have potentially the most to gain from scenario thinking, given their scope for errors resulting from structural changes and uncertainty. If the time and effort are available, then recent augmentations of the basic IL method can be usefully applied to develop and enhance decision makers' forecasting through scenario thinking. For example, evaluation of the actions of powerful stakeholders as scenario storylines unfold (Cairns and Wright, 2018, chapter 3); nuanced analysis of causation within the clusters (Cairns and Wright, 2018, chapter 7)); and use of branching scenarios to stimulate action in scenario team participants (Cairns and Wright, 2018, chapter 6).

By contrast, important trends (i.e., predetermined events) identified by forecasters can be easily assimilated into scenario development since such driving forces are already an identified component of the staged IL scenario development process (see Stage 2 of the basic IL process described earlier).

As we have discussed, scenarios are potentially of most use in forecasting when the scenarios present plausible extreme world contexts for particular focal forecasts. Since appropriate extreme scenarios may or may not emerge from the open-ended basic scenario development process, we advocated using the Backward Logics method as an alternative method to directly identify targeted extreme scenario worlds to underpin the subsequent assessment of forecast bounds. However, future research will be needed to investigate methods that will allow the forecaster to focus on each extreme world in isolation so that the presence of an irrelevant opposing scenario does not contaminate the assessment of a given bound, as the extant laboratory evidence currently suggests it might.

This last point reflects that further work, in general, is needed to establish more effective ways of integrating scenario planning and forecasting. The value of such integration is more easily assessed for forecasts, which can be compared with outcomes, than for scenarios, and this poses a challenge for behavioural researchers in the future. This value may depend, at least to some extent, on the nature of the scenario construction process. As Schirrmeister et al., (2020) discuss, many of the biases associated with judgmental forecasts, such as overconfidence and availability bias, can also be present in scenario planning. Crawford and Wright (2022) provide a common yardstick for the evaluation of: (i) the scenario development process, and (ii) the content of scenarios that are produced by the full variety of extant scenario development methods. These quality issues are reflected in our prior discussion of PESTEL (i.e., full versus abbreviated scenarios, FS/AS) and our later discussion of useful enhancements to the IL method.

Finally, it is important to note that forecasters are not made for their own sake -they are made to support decisions. The integration of scenarios and forecasts has the potential to increase the richness of issues addressed in decision processes, and to draw attention to contextual uncertainties and potential change in these - factors that are often neglected in prescriptive decision analysis (Wright and Goodwin, 1999).

Appendix: The Intuitive Scenario Development Process

In this approach, multiple scenarios are constructed with the aim of bounding the uncertainties faced by an organization (Cairns and Wright, 2018). Each scenario is a consistent and plausible pen picture providing a narrative and multidimensional account of how and why a future might unfold from the current point in time to the horizon year. Importantly, it is the multidimensional space between the scenarios (of which there are usually four developed when applying the IL method) that serves to represent and highlight the degree of uncertainty faced. The greater this space, the greater the uncertainty.

Table A1 gives the sequential stages of the IL scenario development process. The fundamental basis of the scenarios is a judgmental assessment of the "driving forces" (df) that will shape the future. These contextual forces are elicited from workshop participants at the level of the broad PESTEL dimensions – i.e., political, economic, social, technological, environmental, and legal/regulatory. The driving forces will next be categorized as "critical uncertainties" or, alternatively, key "pre-determined trends". For an example of a pre-determined trend, consider that in most of Europe, the proportion of people in the older age bands is becoming larger over time. There is little uncertainty about this trend. By contrast, the outcome of a political election will often be judged as uncertain but with a clear set of plausible outcomes as the result of the candidates or political parties that are offered as voting options.

*** Insert Table A1 about here ***

The next step in the scenario development process is to "cluster" the identified driving forces using "arrows of influence" to link driving forces together within a cluster. An arrow of influence illustrates both time precedence - since whatever driving force starts an arrow must come earlier in time than the driving force that is seen to be causally influenced by that driving force – and so, necessarily, be at the endpoint of an arrow.

Figure A1 gives an example of a cluster relating to the extent to which English market towns will be popular as residences in the future. Here, a cluster of driving forces (some political, some technological, some environmental, some social etc.) are linked together by arrows of influence. Note that only two of the driving forces in the cluster are seen to be pre-determined. The rest of the driving forces are seen as uncertainties – illustrated by use of the words "extent of…", "degree of …", etc. In the example, market towns are survivors from the era of pre-industrialization, a time when there was extensive farming in the UK. Each town has a market square where the farmers in the surrounding farms meet with potential buyers of their meat and vegetables. Once a deal was done the produce was sent by railway to the growing cities. With industrialization, the market towns faded in popularity, with most living in the cities. Could that exodus be reversed in the future? The cluster's content analyses that issue.

*** Insert Figure A1 about here ***

Most driving forces in any cluster will be uncertainties since most of the future is viewed as uncertain, but some will be seen as pre-determined, like demographic change. The next step is the rank-ordering of the formed clusters in terms of their impact on the issue of concern (often the viability of an organization in the future) and also the degree to which the set(s) of out-turns that plausibly could occur when the cluster resolves itself as outcomes are viewed as multiple (i.e. contain uncertainty) or singular (i.e. the cluster's resolution as a set of outcomes is seen as pre-determined). The two highest-impact, most uncertain cluster contents are next chosen as the scenario dimensions.

In Stage 7 of the IL method, scenario storylines are developed where the uncertainties in a focal cluster are resolved as sequences of outcome resolutions. The typical approach is to initiate the flow of event outcomes with an extreme but still plausible resolution. For example, in Figure A1, a new battery type would increase the range of personal electric transport five-fold. By contrast, battery capabilities may increase only marginally. These two "starting outcomes" would thus generate two separate streams of events within the two sets of resolved cluster outcomes. Note that subsequent event resolutions within a particular set of event outcomes must be viewed as a plausible storyline sequence rather than an unconnected set of implausible extreme resolutions. See Cairns and Wright (2018) for more detail on this issue.

In short, scenarios are chosen in the IL scenario development method because they permutate plausible cluster out-turns that are judged to have a high degree of impact on the focal issue of concern. This issue may be as broad as the future of a company or an entire nation (Cairns and Wright, 2018).

References

Alcamo, J. 2001. Scenarios as a tool for international environmental assessments. *Environmental Issue Review* 24. Copenhagen: European Environmental Agency.

Beach, L.R. (2021), Scenarios as narratives. *Futures & Foresight Science*, 3, e58, https://doi.org/10.1002/ffo2.58.

Bennett, E., & Zurek, M. (2006). Integrating epistemologies through scenarios. In W.V Reid, F Berkes, T Wilbanks, and D Capistrano (Eds.) *Bridging Scales and Knowledge Systems: Concepts and Applications in Ecosystem assessment*, Washington: Island Press, 275-294.

Bolger, F., & Harvey, N. (1993). Context-sensitive heuristics in statistical reasoning. The *Quarterly Journal of Experimental Psychology Section A*, 46(4), 779-811.

Bradfield, R., Derbyshire, J., & Wright, G. (2016). The critical role of history in scenario thinking: Augmenting causal analysis within the intuitive logics scenario development methodology. *Futures*, 77, 56-66.

Bulley, A., Redshaw,J., & Suddendorf, T. (2022). *The Invention of Tomorrow: A natural*

History of Foresight. New York: Basic Books.

Bunn, D. W., & Salo, A. A. (1993). Forecasting with scenarios. *European Journal of Operational Research*, 68(3), 291-303.

Cairns G and Wright G (2018). *Scenario Thinking: Preparing Your Organization for the Future in an Unpredictable World*. Second edition. London: Palgrave Macmillan

Chang, W., Chen, E., Mellers, B., & Tetlock, P. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. *Judgment and Decision Making*, 11(5), 509-526.

Chien, Y. W., Wegener, D. T., Hsiao, C. C., & Petty, R. E. (2010). Dimensional range overlap and context effects in consumer judgments. *Journal of Consumer Research, 37(3), 530–542.*

Clements, M. P., & Hendry, D. F. (Eds.). (2002). *A Companion to Economic Forecasting*. Oxford: Blackwell.

Crawford, M.M., and Wright, G. (2022). The value of mass-produced COVID-19 scenarios: a quality evaluation of development process and scenario content. *Technological Forecasting and Social Change*, 183, 121937.

De Baets, S., Önkal, D., & Ahmed, W. (2022). Do risky scenarios affect forecasts of savings and expenses? *Forecasting*, 4(1), 307-334.

Derbyshire, J. (2017). Potential surprise theory as a theoretical foundation for scenario planning. *Technological Forecasting and Social Change*, 124, 77-87.

Derbyshire, J., & Wright, G. (2017). Augmenting the intuitive logics scenario planning method for a more comprehensive analysis of causation. *International Journal of Forecasting*, **33**(1), 254–266.

Döll, P., & Vassolo, S. (2004). Global-scale vs. regional-scale scenario assumptions: implications for estimating future water withdrawals in the Elbe River basin. *Regional Environmental Change*, *4*, 169-181.

Durbach, I. (2019). Scenario planning in the analytic al hierarchy process. *Futures & Foresight Science*, 1 (2), DOI: 10.1002/ffo2.16

Fahimnia, B., Arvan, M., Tan, T., & Siemsen, E. (2023). A hidden anchor: The influence of service levels on demand forecasts. *Journal of Operations Management*, 69(5), 856-871.

Ferretti, V., Montibeller, G., & von Winterfeldt, D. (2023). Testing the effectiveness of debiasing techniques to reduce overprecision in the elicitation of subjective continuous probability distributions. *European Journal of Operational Research*, 304(2), 661-675.

Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces,* 37(6), 570-576.

Fildes, R., & Goodwin, P. (2021). Stability in the inefficient use of forecasting systems: A case study in a supply chain company. *International Journal of Forecasting, 37*(2), 1031-1046.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1), 3-23.

Fildes, R., Goodwin, P., & Önkal, D. (2019). Use and misuse of information in supply chain forecasting of promotion effects. *International Journal of Forecasting*, 35(1), 144-156.

Fildes, R., & Petropoulos, F. (2015). Improving forecast quality in practice. *Foresight: The International Journal of Applied Forecasting*, 36, 5-12.

Fildes, R., & Stekler, H. (2002). The state of macroeconomic forecasting. *Journal of Macroeconomics*, 24(4), 435-468.

Franses, P. H., & Legerstee, R. (2010). Do experts' adjustments on model-based SKU-level forecasts improve forecast quality? *Journal of Forecasting*, 29(3), 331-340.

Glaser, M., Langer, T., & Weber, M. (2013). True overconfidence in interval estimates: Evidence based on a new measure of miscalibration. *Journal of Behavioral Decision Making*, 26(5), 405-417.

Goodwin, P. (2000). Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting*, 16(1), 85-99.

Goodwin, P. (2011). High on complexity, low on evidence: Are advanced forecasting methods always as good as they seem? *Foresight: the International Journal of Applied Forecasting*, 23, 10-12.

Goodwin, P. (2019). Supporting multiattribute decisions in scenario planning using a simple method based on ranks. *Futures & Foresight Science*, 1 (3-4), DOI: 10.1002/ffo2.18

Goodwin, P. (2023). Should we always use forecasts when facing the future? *Foresight: The International Journal of Applied Forecasting*, (69), 20-22.

Goodwin, P., & Fildes, R. (2022). Forecasting with judgment. In: J. Boylan & S. Salhi (Eds.) *The Palgrave Handbook of Operations Research: Recent Perspectives.* London: Palgrave Macmillan, pp. 541-563.

Goodwin, P., Gönül, M. S., & Önkal, D. (2019)a. When providing optimistic and pessimistic scenarios can be detrimental to judgmental demand forecasts and production decisions. *European Journal of Operational Research*, 273(3), 992-1004.

Goodwin, P., Gönül, S., Önkal, D., Kocabıyıkoğlu, A., & Göğüş, C. I. (2019)b. Contrast effects in judgmental forecasting when assessing the implications of worst and best case scenarios. *Journal of Behavioral Decision Making*, 32(5), 536-549.

Goodwin, P., Hoover, J., Makridakis, S., Petropoulos, F., & Tashman, L. (2023). Business forecasting methods: Impressive advances, lagging implementation. *Plos One*, 18(12), e0295693.

Goodwin, P., & Wright, G. (2001). Enhancing strategy evaluation in scenario planning: a role for decision analysis. *Journal of Management Studi*es, 38(1), 1-16.

Goodwin, P., & Wright, G. (2014). *Decision Analysis for Management Judgment*. 5th Edition, Chichester: John Wiley & Sons. Chapter 15.

Green, K. C., & Armstrong, J. S. (2007). Structured analogies for forecasting. *International Journal of Forecasting*, 23(3), 365-376.

Gregory, W. L., & Duran, A. (2001). Scenarios and acceptance of forecasts. In: J.S. Armstrong (Ed.) *Principles of Forecasting: A Handbook for Researchers and Practitioners*, 519-540.

Hall, C. C., Ariss, L., & Todorov, A. (2007). The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organizational Behavior and Human Decision Processes, 103*(2), 277-290.

Hogarth, R. M. (1987) *Judgement and Choice*. 2nd edition, Chichester, England: Wiley.

Jogulu, U. D., & Pansiri, J. (2011). Mixed methods: A research design for management doctoral dissertations. *Management Research Review*, *34*(6), 687-701.

Karelse, J. (2021). Mitigating Unconscious Bias in Forecasting. *Foresight: The International Journal of Applied Forecasting*, (61), 5-14.

Kishita, Y., McLellan, B. C., Giurco, D., Aoki, K., Yoshizawa, G., & Handoh, I. C. (2017). Designing backcasting scenarios for resilient energy futures. *Technological Forecasting and Social Change*, 124, 114-125.

Koutsandreas, D., Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2022). On the selection of forecasting accuracy measures. *Journal of the Operational Research Society*, 73(5), 937-954.

Kunc ,M. (2024) Integrating systems dynamics and scenarios: A framework based on experience. *Futures & Foresight Science*, 6, e174, **https://doi.org/10.1002/ffo2.174**

Lawrence, M., & O'Connor, M. (1992). Exploring judgemental forecasting. *International Journal of Forecasting*, 8(1), 15-26.

MacGregor, D.G. (2001). Decomposition for judgmental forecasting and estimation. In: Armstrong, J.S. (eds) *Principles of Forecasting*. International Series in Operations Research & Management Science, vol 30. Springer, Boston, MA. https://doi.org/10.1007/978-0-306-47630-3_6

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9(4), 527-529.

Millennium Ecosystem Assessment (MA). (2003). *Ecosystems and Human Well-being: A Framework for Assessment*. Washington, DC: Island Press

Morwitz, V. G., & Munz, K. P. (2021). Intentions. *Consumer Psychology Review*, 4(1), 26-41.

National Grid (2023). *Future Energy Scenarios*. https://www.nationalgrideso.com /document/283101/download

Niu, X. (2023). *Judgmental Forecasting: Factors Affecting Lay People's Expectations of Inflation*. Doctoral dissertation, University College London.

O'Brien, F. A. (2004). Scenario planning—lessons for practice from teaching and learning. *European Journal of Operational Research*, 152(3), 709-722.

Önkal, D., & Gönül, M. S. (2005). Judgmental adjustment: A challenge for providers and users of forecasts. *Foresight: The International Journal of Applied Forecasting*, (1), 13-17.

Önkal, D., Sayım, K. Z., & Gönül, M. S. (2013). Scenarios as channels of forecast advice. *Technological Forecasting and Social Change*, 80(4), 772-788.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, 29(3), 261-265.

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, 15, 353–375.

Rowland, N.J. & Spaniol, M.J. (2022). The strategic conversation, 25 years later: A retrospective review of Kees van der Heijden's Scenarios: The Art of Strategic Conversation, *Futures & Foresight Science*, https://doi.org/10.1002/ffo2.102

Schirrmeister, E., Göhring, A. L., & Warnke, P. (2020). Psychological biases and heuristics in the context of foresight and scenario processes. *Futures & Foresight Science*, *2*(2), e31.

Schnaars, S. P., & Topol, M. T. (1987). The use of multiple scenarios in sales forecasting: An empirical test. *International Journal of Forecasting*, *3*(3-4), 405-419.

Schoemaker, P. J.H. (1991). When and how to use scenario planning: a heuristic approach with illustration. *Journal of Forecasting*, 10(6), 549-564.

Schoemaker, P.J.H. (1993). Multiple scenario development: its conceptual and behavioral foundations. *Strategic Management Journal*,14, 193-123.

Scoblic, J. P., & Tetlock, P. E. (2020). A better crystal ball: The right way to think about the future. *Foreign Affairs*, 99, 10.

Sherif, M., Taub, D., & Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology*, 55(2), 150-155.

Spaniol, M. J., & Rowland, N. J. (2019). Defining scenario. *Futures & Foresight Science*, 1(1), e3.

Spaniol, M. J., & Rowland, N. J. (2023). AI-assisted scenario generation for strategic planning. *Futures & Foresight Science*, e148.

Sjöberg, L. (1982). Aided and unaided decision making: Improving intuitive judgement. *Journal of Forecasting*, 1(4), 349-363.

Taylor, P. F., & Thomas, M. E. (1982). Short term forecasting: horses for courses. *Journal of the Operational Research Society*, *33*(8), 685-694.

Taleb, N. (2007). *The Black Swan: The Impact of the Highly Improbable*. London: Allen Lane., pp112.

Tetlock, P. E. (2017). *Expert Political Judgment*. Princeton: Princeton University Press.

Tetlock, P. E., & Gardner, D. (2016). *Superforecasting: The Art and Science of Prediction*. New York: Random House.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293-315.

Van der Heijden, K. (2005). *Scenarios: The Art of Strategic Conversation*. Chichester: John Wiley & Sons.

Wright, G., Cairns, G., & Goodwin, P. (2009). Teaching scenario planning: lessons from practice in academe and business. *European Journal of Operational Research*, 194(1), 323-335.

Wright, G., & Goodwin, P. (1999) Future-focussed thinking: combining scenario planning with decision analysis. *Journal of Multi-Criteria Decision Analysis*, 8(6), 311-321.

Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: Enhancing the scenario method. *International Journal of Forecasting*, 25(4), 813-825.

Intuitive Logics scenarios	Forecasts
• Narratives.	• Numbers.
• Rationale/causation provided and documented.	• Rationale often not made explicit (though may be In Delphi).
• Natural way of thinking about future and therefore persuasive.	• Less natural.
• Only estimates bounds.	• Most commonly aims for modal or mean outcome.
• Aims to span uncertainty.	• Does not always aim to span uncertainty.
• Probabilities not estimated, but danger of scenario bias.	• Amenable to probabilities and hence biases.
• Multivariate descriptions of future.	• Single variable estimation of future outcome.
• Always judgment-based.	• Both judgment & algorithm-based.
• Impact of factors determines scenario formulation.	• Impact of outcomes not considered.
• Group-based.	• Can be group, individual or algorithm-based.
• Under-developed theoretical underpinnings.	• Theoretical underpinnings when algorithm-based.
• Long term focus.	• Short to medium term focus.
• Difficult to verify validity.	• Some, but not perfect, scope for assessing validity.
• Iterative judgmental process.	• Algorithm & subsequent judgmental adjustment is common process.
• Focus is present time to future.	• Time series focus is from past history to future.
• Subsequent focus on robust options.	• Subsequent focus on 'best' options.
• Always involves judgmental decomposition.	• In practice, rarely involves structured decomposition.

Table 1 Narrative scenarios vs forecasting as widely practiced

Effects of given (received) scenarios on forecast adjustments made to one-period-ahead point forecasts within 18 time-series plots (Onkal et al, 2013).The scenarios utilized were "worst-case" and "best-case" e.g: "*This is one of our best performing mobile phones with complex features…it was a hit product when first introduced.. With new features added from time to time…price reasonably positioned …desirable object ..teenager consumer group .. Future looks bright…expected to carry on its upward trend…"*
Effect of providing pessimistic and optimistic scenarios on judgmental demand forecasts and production decisions (Goodwin et al., 2019a) Exemplar optimistic scenario: *"Product K, a mobile phone ... has stable demand ... combination of necessary features to compete successfully ... nicely designed ... fairly positioned price ... receives positive comments in industry magazines ... good feedback from customers ... given the recent economic conditions ... there may be a slightly higher demand for this product in periods to come."*
The implications of presenting worst and best case scenarios on judgmental forecasts (Goodwin et al., 2019b). Exemplar best case scenario: *Product A has an expanding and loyal user base. It is improved with some significant and attractive modifications ... receive enthusiastic reviews ... the product continues to be regarded as the best of its type ... No models directly compete ... the product's popularity is very high and it is likely to attract a significant number of new purchasers"*
Scenarios affecting forecasts of savings and expenses (De Baets et al., 2022). Two received scenario contexts, "unexpected–expense" and "unexpected income-loss", e.g: "*Imagine ..you come home after a busy day ...you open your door and the hallway is full of water. A water pipe has broken ... you hurry to shut off the water supply ... You call a plumber. The quote he gives amounts to 80% of your monthly income. How does this affect your savings ... forecasts for the next three months?"*

Table 2: Laboratory scenarios as contexts for judgmental forecasts

Stage	Activity
Stage 1: Setting the scenario agenda	Defining the focal issue of concern and process, and setting the scenario timescale.
Stage 2: Determining the driving forces	Eliciting a multiplicity of wide-ranging driving forces. utilizing the PESTEL dimensions
Stage 3: Clustering the driving forces	Clustering causally related driving forces, testing and naming the clusters.
Stage 4: Using the Impact/uncertainty matrix	Ranking each of the clusters to determine their relative impact on the issue of concern and also the relative degree of uncertainty as to their resolutions as a cluster-based set of outcomes.
Stage 5: Identifying the scenario dimensions	Selecting two clusters - each with (i) a high degree of impact on the focal issue of concern and (ii) a high degree of uncertainty about the out-turn of the cluster's contents as a set of outcomes.
Stage 6: Scoping the scenarios	Building a broad set of descriptors for each of the four scenarios.
Stage 7: Developing the scenario storylines	Developing scenario storylines, including key events and their chronological structure, i.e., the 'why' of what happens.

Table A1: The Basic Intuitive Logics Standard Approach to Scenario Development (adapted from Cairns and Wright, 2018)

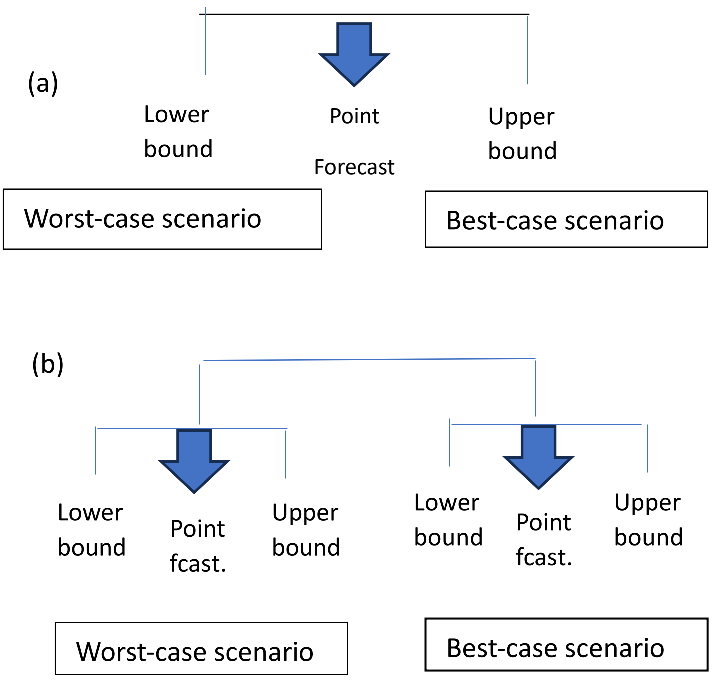

Figure 1. First and second order uncertainty representation

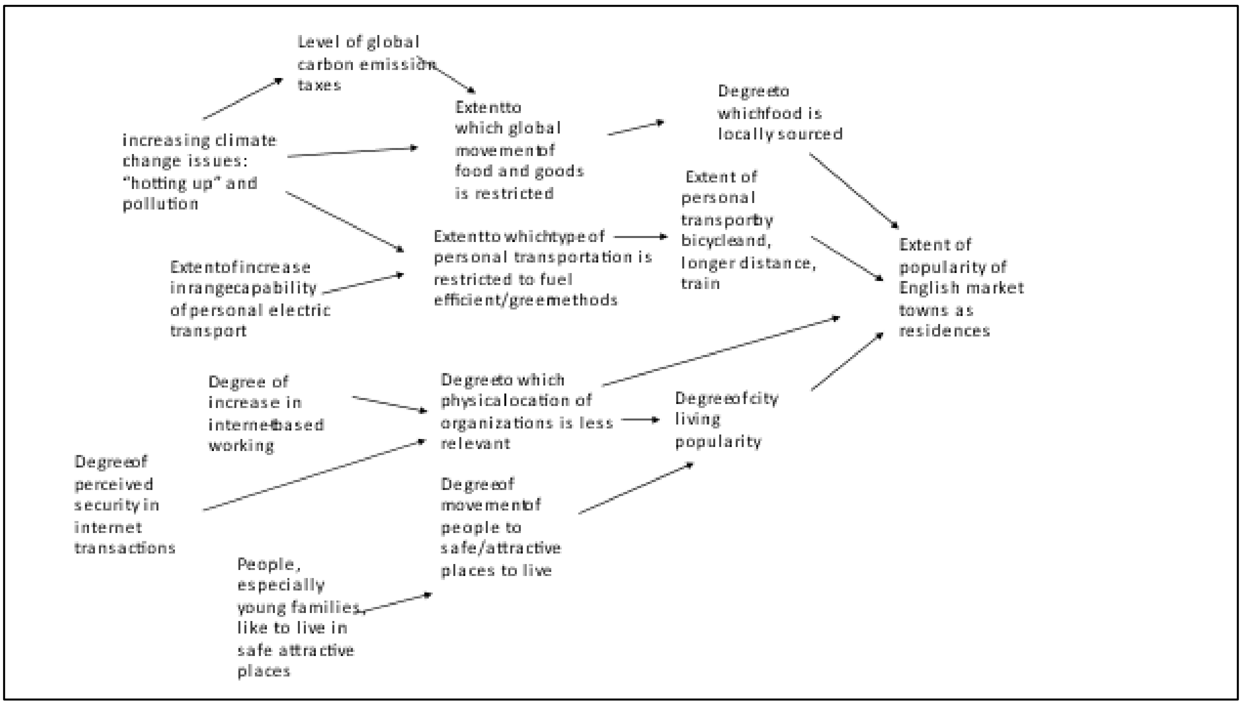

Figure A1 An example cluster of driving forces