Contrast effects in judgmental forecasting when assessing the implications of worst- and best-case scenarios

Paul Goodwin
School of Management
University of Bath
BA2 7AY
United Kingdom

M. Sinan Gönül,
Newcastle Business School
Northumbria University
United Kingdom

Dilek Önkal
Newcastle Business School
Northumbria University
United Kingdom

Ayse Kocabıyıkoğlu
School of Management
Sabanci University
Turkey

Itir Göğüş
Faculty of Business Administration
Bilkent Unibersity
Turkey

Contrast effects in judgmental forecasting when assessing the implications of worst- and best-case scenarios

Abstract

Two experiments investigated whether individuals' forecasts of the demand for products and a stock market index assuming a best or worst-case scenario depend on whether they have seen a single scenario in isolation or whether they have also seen a second scenario presenting an opposing view of the future. Normatively, scenarios should be regarded as belonging to different plausible future worlds so that the judged implications of one scenario should not be affected when other scenarios are available. However, the results provided evidence of contrast effects in that the presentation of a second 'opposite' scenario led to more extreme forecasts consistent with the polarity of the original scenario. In addition, people were more confident about their forecasts based on a given scenario when two opposing scenarios were available. We examine the implications of our findings for the elicitation of point forecasts and judgmental prediction intervals and the biases that are often associated with them.

Keywords: Demand forecasting; Stock market forecasting; Best and worst case scenarios; Prediction intervals; Context effects; Contrast effects.

Contrast effects in judgmental forecasting when assessing the implications of worst- and best-case scenarios

1. Introduction

When considering the future, people often speak in terms of 'best-case' and 'worst-case' scenarios. Scenarios are also used in businesses and other organisations to plan strategies that will cope with alternative possible futures (e.g. Goodwin and Wright, 2014). A key characteristic of these scenarios is that they represent mutually exclusive (though usually not mutually exhaustive) plausible combinations of future events. Each scenario represents a different possible future world with a specific set of implications for a decision maker. For example, a scenario involving a series of political events that result in peace in the Middle East could imply a low price for oil on the world markets at a given future date. A scenario involving continued conflict in that region could imply a high price for oil, as extraction is suspended and supplies are disrupted. Note that the implications are conditional only on the occurrence of the scenario on which they are based. The possibility that the 'continued conflict' scenario might prevail should have no relevance to the estimated price of oil that would be associated with the 'peace' scenario -each scenario represents a different future world.

However, it is known that effect of a stimulus is often context dependent (Plous, 1993, p38). As Plous argues: '[People] do not perceive and remember material in isolation; they interpret new information in light of past experience and the context in which the material occurs.' This suggests there is a possibility that, when presented with different mutually-exclusive scenarios, and asked to estimate the implication of each scenario, people's estimate for a given scenario may be influenced by the content of the alternative scenarios that they have recently studied. For example, having just read a worst-case scenario, a person's point forecast of a variable, assuming a best-case scenario may be more or less optimistic than it would be if the latter scenario had been considered in isolation. If it was more optimistic, it would represent a contrast effect –ironically, access to the worst-case scenario would have had the effect of increasing optimism. If it was more pessimistic, it would represent an assimilation effect –the information in the worst case scenario would have been assimilated with that in the best case scenario and led to a damping of the forecaster's optimism. For the reasons we present below, we have reasons to expect that contrast effects will occur when point forecasts are based on extreme scenarios. In cases where organizations are employing scenario planning tools, such a bias could have important implications for the quality of strategic decisions. However, to our knowledge, this possibility has not previously been investigated in the literature.

Forecasts are also often presented in the form of prediction intervals. Typically, the widths of judgmental prediction intervals are too narrow, suggesting that people are overconfident that an elicited interval will capture the realised outcome (e.g. Ben-David, Graham and Harvey, 2013). Confidence in one's

forecasts tends to increase disproportionately as the amount of available information increases, even when this information is irrelevant to the forecast (Oskamp, 1965; Hall, Ariss and Todorov, 2007). Given that information in one scenario is not relevant to a forecast based on a second scenario, we also examine whether the implied level of confidence associated with such a forecast increases when the opposing scenario is also available.

In this paper we present two experiments that were used to explore whether the perceived implications of worst and best-case scenarios, expressed as point forecasts and prediction intervals, are context dependent, and whether any perceived dependence between mutually-exclusive scenarios is related to the 'extremity' of the arguments presented within them. Following a review of the relevant literature, we give details of the design and implementation of the experiments, before presenting the results and analysis and discussing their practical implications.

2. Literature review

2.1 Examples of context effects

Perhaps the best known example of context dependence is the halo effect, which was discovered in the 1920s (Thorndike, 1920). In a task that involved rating army officers on a number of attributes, such as intelligence, physique and leadership, Thorndike found that people were unable to rate each attribute independently of the others. For example, there was evidence that an officer who was perceived to be intelligent also tended to be perceived as having a strong physique. Although there were problems with Thorndike's measurement scales a large number of subsequent studies have reached similar conclusions in a wide range of domains (e.g. Nisbett and Wilson, 1977; O'Donnell and Schultz, 2005; Chernev and Blair, 2015). Context dependence has also been found to apply in choice tasks where the addition to the choice set of alternatives that are vastly inferior to existing options -and hence irrelevant to the choice between them - changes the way that these existing options are valued (Louie, Khaw and Glimcher, 2013; Soltani,De Martino and Camerer., 2012; Vlaev, Chater, Stewart and Brown, 2011). Context effects on choice can also be found where consumers are presented with superior products or services that are unavailable –so called 'phantom decoys' (Scarpi and Pizzi, 2013, Trueblood and Pettibone, 2017). For example, these can be influential when consumers are belatedly informed that a favoured product is out of stock (Pizzi and Scarpi, 2013). In other work, Stewart, Chater and Brown (2006) suggest that the values people attach to the attributes of an option in decision making can depend on other values in the decision context, such as the attribute values of competing options. Similarly, the provision to judges of potential anchors, which have values that are so implausible that they are irrelevant to the estimation task in hand, has been found to influence the subsequent values estimated. For example, in one study people who were initially asked whether Gandhi's age at death was higher or lower than 9, produced significantly lower estimates of his age

than those who were initially asked where his age at death was higher or lower than 140 (Strack and Mussweiler, 1997).

2.2 Assimilation and contrast effects

Context dependence can result in either assimilation effects or contrast effects. An assimilation effect occurs when a person's response to a target stimulus is positively correlated with contextual information, as in the anchoring bias described above that is widely observed in forecasting tasks (Lawrence and O'Connor, 1992; Bolger and Harvey, 1993). This effect can also be seen in psychology experiments where tasks completed earlier set a context for later tasks which the experimenter has intended to be viewed independently. For example, in an experiment-based study of judgmental extrapolation of time series, Harvey and Reimers (2013) partly attributed their finding that forecasters tended to damp trends to an assimilation effect. In this case, it appeared that people tended to regress their estimate of a given trend towards the mean trend that they had observed in other series used in the experiment.

When a person's response to a target stimulus is negatively correlated with contextual information, a contrast effect occurs. For example, in an early experiment, participants rated weights as being lighter than they actually were, immediately after lifting a heavier weight (Sherif, Taub and Hovland, 1958). Contrast effects have been found in areas ranging from legal decision making (Kelman, Rottenstreich and Tversky, 1996) to consumer choice, where a product may appear to be less attractive than it would be when considered in isolation, when it is compared to much more attractive alternatives (Simonson and Tversky, 1992).

2.3 Factors determining whether an assimilation or a contrast effect occurs

A number of factors may be responsible for whether an assimilation or contrast effect occurs. Basing their theory on the inclusion/exclusion model (IEM) model of Schwarz and Bless (1992a), Förster, Liberman and Kuschel (2008) suggest that, when similarity-orientated processing is invoked, this is likely to result in assimilation. In this mode of thinking a person will focus on the similarities between the context and the target stimuli and they may base their interpretation of the target on the characteristics of the context. Dissimilarity-orientated, or comparative, processing on the other hand is likely to be associated with a contrast effect. In this case, the focus will be on the differences between the context and target. This raises the question of which features of a judgmental task are likely to lead to each type of thinking. The literature has identified the following features.

(i) *The distance between context and target*. Early research found that, when people made quantitative judgments, the distance on a scale between the context and the target determined the type of response. In an experiment, where people had to assess how heavy pairs of weights were, the first weight (the context) was defined as the one at the top of the scale and the second,

generally lighter weight, had to be rated against it (Sherif et al., 1958). If the two weights were similar the estimate tended to move upwards towards the 'context' weight (an assimilation effect), but if the 'context' weight was much heavier than the target a contrast effect occurred and the second weight was regarded as being lighter. The contrast effect increased the greater the distance of the context weight from the target weight. More recent research by Chien, Wegener, Hsiao and Petty (2010) suggests that, where the plausible ranges of values of the context and target stimuli are perceived to overlap, an assimilation effect will probably occur. When they do not, a contrast effect is likely.

(ii) *The ambiguity of target stimulus.* Herr, Sherman and Fazio (1983) showed that an assimilation effect is likely when the target stimulus is both ambiguous and in close proximity to the contextual stimulus. Ambiguity is the extent to which a stimulus can be interpreted in diverse ways. When a target stimulus is highly ambiguous the contextual stimulus may prime the target to be interpreted in a way that is consistent with it, thereby leading to assimilation. However, an unambiguous target stimulus would, according to Herr et al. be expected to result in a contrast effect, especially where both the target and contextual stimuli are extremes.

(iii) *The distinctiveness of the contextual information.* Stapel and Winkielman (1998) found that, where the contextual information is distinct and accessible, it is more likely to act as comparison standard and hence lead to a contrast effect. Indistinct or abstract contextual information is less accessible and hence likely to lead to such an effect.

(iv) *The dimensional relevance of the contextual information.* The contextual information will be more likely to lead to a comparative mode of thinking and hence to a contrast effect when is seen as being relevant to the dimension on which the target stimulus is being assessed (Stapel and Winkielman,1998). For example, contextual information that specifies a how heavy a weight is will clearly be relevant when the heaviness of another weight is being assessed. Contextual information about the shape of the weight would not have dimensional relevance.

(v) *Whether an explicit evaluation is made of the contextual information.* An explicit evaluation of the context is likely to make its mental representation concrete and distinct. This reification means that the contextual information is likely to form a comparison standard and hence be conducive to a contrast effect (Stapel and Winkielman,1998).

(vi) *The time interval between presentation of the context and target stimuli.* Manis and Moore (1978) found that short time intervals between the presentation of context and target stimuli (these were polarised messages concerning social issues or irrelevant messages) tended to lead to contrast effects while assimilation effects were associated with longer intervals.

We are not aware of any research that has looked at the possibility that one scenario can act as a context for another when people are judging the implications of a given scenario in order to make a forecast. However, if the findings of the foregoing research apply to scenarios, then a number of inferences can be drawn. First, if the context and target scenarios are extremes (as in 'best-case and 'worst-case' scenarios), then this is likely to lead to a contrast effect. However, the extent of this may depend on the degree of extremity of the context scenario. Secondly, the probability of a contrast effect will depend on the extent to which the scenarios lack ambiguity and are distinctive. For example, scenarios that convey a mixture of optimism and pessimism may be perceived as being more ambiguous and hence more conducive to an assimilation effect. An assimilation effect may also be more likely when the scenarios describe the future on different dimensions. For example, if a context scenario relating to a country confines itself to describing future rate of inflation and unemployment in a country, while the target scenario only describes future rates of pollution and crime, the two scenarios can be 'interweaved' and hence assimilated. None of the variables described in the context scenario would provide a comparison standard for those in the target scenario. Finally, we anticipate that the act of making a forecast based on the context scenario (in addition to a forecast for the target scenario) will reify the scenario and hence make it more likely to act as a basis of comparison and engender a contrast effect.

2.4 Hypotheses

This discussion suggests the following hypotheses when people undertake a task of making point forecasts based on worst and best case scenarios that are extreme, unambiguous, non-abstract and describe possible futures based on the same dimensions.

H1: Exposure to a worst case scenario will lead to more optimistic forecasts based on a best-case scenario. Similarly, exposure to a best case scenario will lead to more pessimistic forecasts based on a worst-case scenario (i.e. contrast effects will be observed).

H2: The degree of extremity of the context scenario will determine the size of the contrast effect.

H3: The occurrence of a contrast effect is dependent on whether or not a forecast was previously made for the context scenario.

Point forecasts do not provide any information about the degree of uncertainty associated with a forecast and this information can be crucial for any subsequent decisions (Goodwin, 2014). Prediction intervals do provide this information, but as indicated earlier, when they are based on judgment, there is a tendency for them to be too narrow. As a result, they manifest overconfidence in that the probability of them capturing the outcome tends to be less than the stated coverage probability (e.g. Ben-David et al., 2013. Welsh and Begg, 2018). Given the evidence that overconfidence increases when more information is available, even when the extra information is irrelevant (e.g. Oskamp, 1965; Hall et al., 2007), we expect

that people will produce narrower prediction intervals for forecasts based on a given scenario when they also have access to information on an opposing scenario. Hence, we have the following hypothesis.

H4: Prediction intervals produced for a given scenario are narrower when an opposing scenario has also been presented to the forecaster.

3. Method: Experiment 1

3.1 Design of experiment

Experiment 1 was designed to test H1, H2 and H4. The participants' task related to the production of demand forecasts based on the assumption that a given 'background' scenario would prevail.[1] A total of 114 business students from Bilkent and Sabancı Universities participated in a task that had a paper-and-pencil format. The task involved the use of judgment to produce the forecasts for a company marketing high technology products and services such as fitness monitoring devices, drones and a movie streaming service. The participants received an individual forecast form that included background information about the company and its forecasting procedure, together with time-series plots showing past demand for eight products (see Appendix 1 for an example of an instruction form supplied to a participant for one of the products). Each plot displayed 10 months of data. Participants received these series and their associated scenario in a randomized order. The formula used for the generating each series was:

$$Y_t = 125 + e_t \qquad e_t \sim N(0,20) \qquad (1)$$

where Y_t = the demand in month t and e_t = the level of noise in month t.

Because the focus of the study was on the possible interaction between scenarios, the plots were intended to have, as far as possible, a neutral effect, providing guidance only on the typical level and variation of demand. Hence, they did not contain autocorrelations, trends or seasonal patterns.

The participants were randomly assigned to one of three groups. Those in Group 1 were presented with a single best or worst-case scenario (Depending on the product or service, this scenario was either extreme or moderate in its pessimism or optimism the same level of optimism or pessimism was always associated with a given product). The participants were then asked to assume that this scenario applied to market conditions in month 20 (ten months ahead of the latest demand figure) and to produce: i) a point forecast, ii) a best case forecast and iii) a worst case forecast of demand in this month. This process was carried out for 8 products or services. The extreme or moderate levels of optimism or pessimism were achieved by varying the tone of the scenario vignettes by manipulating the words used (the full set of scenarios are available from the authors). The content and the amount of information conveyed were kept

[1] There is a clear distinction between scenarios and forecasts and the terms are not being conflated here. In the experiment the scenarios provided the context and the forecasts provided the estimates of future demand conditional on that context.

the same across scenarios with different levels of optimism or pessimism. In total, the participants saw two extreme and two moderate worst-case scenarios and two extreme and moderate best-case scenarios (i.e. a total of 8 scenarios, one for each product).

Participants in Group 2 received the same scenarios as those in Group 1, but they also received a moderate second scenario that was opposite in polarity to the first For example, when the first scenario was extremely pessimistic, the second conveyed moderate optimism. Having read the two scenarios, they were then asked to produce a point forecast and a best case and worst case forecast of demand in month 20 assuming that the first scenario was applicable to the market conditions in that month. Following this, they were asked to produce a second set of forecasts, assuming that the second scenario was applicable. The main purpose of asking for the second set of forecasts was to ensure that both scenarios were read by the participants.

Participants in Group 3 carried out the same task as those in Group 2, except that the second scenario was extreme in tone. For example, when the first scenario was extremely pessimistic, the second conveyed extreme optimism. Having made all their forecasts participants completed an exit questionnaire that was designed to ascertain their feelings about the task and the role of the scenarios.

One-way ANOVAs were conducted to compare the mean point forecasts of the groups for cases where the target scenario was i) extremely pessimistic, ii) moderately pessimistic, iii) moderately optimistic and iv) extremely optimistic. It was considered to be reasonable to average the pairs of point forecasts for each condition as the underlying generating mechanism was the same (see equation 1).

In the exit poll the participants were given questions that asked them to indicate on a 1 (Totally disagree) to 5 (Totally agree) scale their evaluation of the scenarios they were given. In particular, they rated whether they believed the scenarios "enhanced their future-focused thinking", "were useful in constructing the forecasts", "were clear to understand", "were realistic" and "provided important additional information that helped in constructing the forecasts". Again, one-way ANOVAs were applied to the responses.

3.2 Results of experiment 1

3.2.1 Point forecasts

Tables 1(a) to 1(d) show the results of the ANOVAs. All indicated significant differences between the groups' mean point forecasts (in all cases the test statistic was $F_{2,111}$; this had values between 3.6 and 19.5 and p-values ranged from 0.03 to below 0.0005). Also, in all cases, apart from the results in Table 1(b), eta-squared suggested that the presence of 'opposing' scenarios tended to have a medium to large effect on the forecasts (Cohen, 1988). The results provide evidence of a significant contrast effect in all cases. For example, those who only received a moderate pessimistic scenario tended to produce higher point forecasts

than those who also received a moderate or extreme optimistic scenario and who were asked to assume that the pessimistic scenario would prevail (see Table 1a). Access to an optimistic scenario appeared to make people more pessimistic when judging the implications of the pessimistic scenario. The reverse was true when people were asked to make point forecasts assuming that an optimistic scenario would ensue. In this case, access to a pessimistic scenario tended to make them more optimistic. In addition stronger 'opposing' scenarios tended to lead to a greater contrast effect, though this effect was relatively small when point forecasts were being made for an extreme optimistic scenario (see Table 2(d)). These results provide support for H1 and H2.

Please insert Tables 1(a) to 1(d) about here

3.2.2 Prediction intervals

As indicated above, participants were also asked to produce best and worst case forecasts under the assumption that a given scenario applied. The width of the interval between these two forecasts can be interpreted as a measure of the degree of uncertainty that the participant associated with their point forecast. Table 2 displays the results of one-way ANOVAs that were used compare, for the three groups, the mean interval widths of forecasts made for the four types of scenario. It can be seen in all cases that the mean intervals became significantly narrower when 'opposing' scenarios were presented (in all cases the test statistic was $F_{2,111}$; this had values between 3.3 and 35.3 and p-values ranged from 0.041 to below 0.0005). They were also narrower when the 'opposing' scenario was extreme rather than moderate. Thus, ironically, participants typically were more confident about their point forecast that assumed a given scenario would prevail when they were also presented with scenarios that described an opposite view of possible future market conditions. Not only was there a contrast effect, but access to opposing scenarios also appeared to reduce the uncertainty that participants had about their point forecasts.

Please insert Table 2 about here

3.2.3 Exit poll

Recall that, in the exit poll, the participants rated whether they believed the scenarios "enhanced their future-focused thinking", "were useful in constructing the forecasts", "were clear to understand", "were realistic" and "provided important additional information that helped in constructing the forecasts". One-way ANOVAs showed, the participants' responses, summarized in Figure 1, were not significantly different across the three groups (in all cases the test statistic was $F_{2,111}$; all *p*-values were greater than 0.4). Furthermore, t-tests revealed that they were significantly different than 3 (all *p*-values<0.001, two-tail, for all comparisons), indicating that the participants found the scenarios to be meaningful and helpful. The exit poll also asked the participants to indicate on a 1 to 5 scale whether they found the products and services of the company attractive and appealing. The average response to this question was 3.6 for those receiving a

single scenario (Group 1), 3.77 for those receiving two scenarios including a moderate opposing scenario (Group 2), and 3.51 for those receiving two scenarios including an extreme opposing scenario (Group 3); all were significantly higher than 3 when t-tests were applied (all *p*-values<0.001, two-tail), suggesting the participants found the task interesting, and were motivated to complete the forecasts.

Please insert Figure 1 about here

4. Method: Experiment 2

4.1 Design of Experiment

Experiment 2 was conducted to investigate: (i) whether the findings of contrast effects in the first experiment also applied in a different context (this was to assess how generalizable the findings might be) and (ii) to test H3, that is to establish whether a contrast effect only applied when a point forecast was also made for the context scenario. Recall that the literature suggests that requiring a point forecast would make the context scenario more concrete in the forecast's mind and hence be more likely to lead to a contrast effect. To further test the robustness of the findings of experiment 1, in this second experiment the forecasts for the target scenario were elicited after the forecasts for the context scenario, rather than before. All the scenarios in this experiment were extreme.

This time the participants' task related to the production of stock market forecasts based on the assumption that a given scenario relating to the economic health of a country would prevail in ten months' time. A total of 90 business students from Bilkent and Sabancı Universities participated in a task that, once again had a paper-and-pencil format. The participants assumed the role of forecasters in an international consulting company. They each received an individual forecast form that included background information about the consultancy company and its forecasting procedure, together with time-series plots showing the values of each country's stock market index for the past ten months. Forecasts for eight countries, presented in random order were required (see Appendix 2 for an example of an instruction form supplied to a participant for one of the countries). The formula used for the generating each series was:

$$I_t = 100 + i_t \qquad i_t \sim N(0,30) \qquad (2)$$

where I_t = the value of the index in month t and e_t = the level of noise in month t.

Again, because the focus of the study was on the possible interaction between scenarios, the plots were intended to have, as far as possible, a neutral effect, providing guidance only on the typical level and variation of the index.

The participants were randomly assigned to one of three groups of 30. Those in Group 1 were presented with a single scenario. Depending on the country. this scenario was either extremely pessimistic or optimistic. The participants were then asked to assume that this scenario applied to economic conditions in month 20 and to produce: i) a rating of the scenario for its degree of optimism or pessimism on a -5 (extreme pessimism) to +5 (extreme optimism) scale, ii) a point forecast, iii) a best case forecast and iv) a worst case forecast of the index's value in this month. In total, each participant received four extreme worst case scenarios and four extreme best-case scenarios.

Participants in Group 2 received the same best and worst-case scenarios as those in Group 1, but they also received a second extreme scenario that was opposite in polarity to the first. . As far as possible, the aspects of each economy that was described in the worst and best case scenarios for a given country were the same, they simply differed in their polarity. For example, when the first scenario was extremely optimistic, the second conveyed extreme pessimism. Having read the two scenarios, they were then asked to give a rating of the degree of optimism or pessimism implied by the scenarios. However, the point, best case and worst case forecasts were only required for the second scenario. This was to establish whether not being required to make a forecast for the context scenario reduced the likelihood of a contrast effect (as suggested by H3). The requirement to rate both scenarios was intended to ensure that both had been read.

Participants in Group 3 carried out the same task as those in Group 2, except that the they were also required to produce forecasts for both scenarios.

Having made all their forecasts participants completed an exit questionnaire that was designed to ascertain their feelings about the task and the role of the scenarios. One-way ANOVAs were conducted to compare the mean point forecasts of the groups for cases where the target scenario was i) extremely pessimistic and ii) extremely optimistic. The exit poll asked the same questions as those in Experiment 1, plus questions on whether the participants had carefully read the provided scenario and whether, overall, they felt satisfied with their forecasts (the responses were again on a scale from 1(totally disagree) to 5 (totally agree). As before, one-way ANOVAs were applied to the responses.

4.2 Results of Experiment 2

4.2.1 Point forecasts

Table 3 show the results of the ANOVAs . Both indicated significant differences between the groups' mean point forecasts ($F_{2,87}$ equalled 13.4 and 7.1, yielding p-values of less than 0.0005 and 0.001, respectively). The eta-squared values suggested that the nature of the task carried out by each group tended to have a medium to large effect on the forecasts (Cohen, 1988). The tasks of Group 1, who received a single scenario, and Group 3 who received two scenarios, including an extreme opposing scenario, largely replicated those in Experiment 1. Forecasts made for worst-case scenarios tended to be more pessimistic when an opposing

best-case scenario had also been read, indicating a contrast effect (p =0.0004 in a one-tail pairwise comparison t-test of the group 1 and Group 3 mean forecasts, which is easily significant after Bonferroni adjustment). Similarly, forecasts made for a best-case scenario tended to more optimistic when an opposing worst-case scenario had also been read –though here the effect was not as large and is not quite significant at the 5% level (p =0.073 in a one-tail pairwise comparison t-test). Despite this, the results broadly provide further support for H1.

Please insert Table 3 about here

However, Group 2, who received two scenarios, including an extreme opposing scenario did not exhibit a contrast effect. Recall that, despite seeing and rating the two scenarios this group, unlike Group 3, only made a forecast for one of them. The forecasts of participants in this group tended to move in the direction of the context scenario, indicating a small assimilation effect (though this was only significant at the 5% level, after Bonferroni adjustment, when forecasts were made for best case scenarios). This finding supports H3. It appears that, for a contrast effect to occur, a forecast also has to be made for the context scenario. This is consistent with the idea that the act of making a forecast based on a scenario reifies that scenario so it is more likely to act as a comparison standard.

4.2.2 Prediction intervals and ratings of scenarios

Table 4 shows that the finding of Experiment 1, that providing two scenarios leads to a narrowing of the mean prediction intervals, was also replicated in the second experiment. The mean interval width for the target scenario of participants in Group 1, who received one scenario, was significantly narrower than that of Group 3, who received two scenarios (p< 0.0005 in one-tail pairwise comparison t-test for both worst-case and best-case target scenarios). This provides further support for H4. However, once again the result for Group 2 ,who received two scenarios but only produced a forecast for one, did not conform to this pattern. Indeed, the mean interval widths for this group were not significantly different at the 5% level from those of Group 1 who only saw the target scenario (p= 0.084 in a one-tail pairwise comparison t-test for worst-case target scenarios; p =0.975 for best-case target scenarios).

Please insert Table 4 about here

Please insert Table 5 about here

Table 5 show the mean ratings made by the participants of the degree of optimism and pessimism of the target and context scenarios. Although they are all in the expected directions, the mean ratings tended to be conservative. Despite the scenarios representing extreme economic conditions, their mean absolute rating was 2.8 (recall that the maximum absolute value was 5 for extreme optimism or pessimism). ANOVAs

indicated that Group 2, who saw two scenarios, but only produced a forecast for one, produced ratings that were significantly more conservative than the other group(s) when the target scenario was extremely pessimistic and the context scenario was extremely optimistic ($F_{2,87}$ =4.1, p=0.02 for the target scenario. $F_{1,58}$ = 5.7, p=0.02 for the context scenario). However, there was no significant difference between the groups when the target scenario was extremely optimistic. Nor were there any significant differences between the mean ratings of the target and context scenarios by the participants who saw both scenarios.

4.2.3 Exit poll

The ANOVAs showed no significant differences between the three groups in their responses to any of the questions (p-values ranged from 0.1 to 0.9). On all questions, the mean response of all 90 of the participants was significantly greater than 3 (p-values<0.001 on two-tail t-tests, for all comparisons), suggesting in particular, that they agreed that the scenarios enhanced their future-focussed thinking, were useful in constructing the forecasts, were clearly understood, were realistic and provided important additional information when the forecasts were constructed (for brevity, we have not included a radar chart for this experiment).

5. Discussion

The experiments yielded four main findings. First, people tended to manifest a contrast effect. When asked to judge the implications of a best-case scenario, individuals made forecasts that were more positive if they also had access to a worst-case scenario. Similarly, when asked to judge the implications of a worst-case scenario, their forecasts were more negative if they also had access to a best-case scenario. Second, as shown by the results of Experiment 1, the extent of the contrast effect was dependent on the extremity of the context scenario –less extreme context scenarios tended to have a smaller contrast effect. Third, the contrast effect was dependent on the participant making a forecast for the context scenario. Fourth, access to the 'opposing' scenario made participants more confident about their point forecasts, but again only if they had made a forecast for the context scenario.

5.1 Contrast effects

The results are therefore consistent with research in other contexts which suggests that a contrast effect will occur when the target scenarios is unambiguous (Herr at al., 1983) and the context scenario is regarded as a standard for comparison because it is extreme (Herr at al., 1983), distinct, rather than abstract, relevant to the dimension of the target stimulus (Stapel and Winkielman, 1998) and reified by being subject to an evaluation. The short time interval between the reading of the context and target scenarios may also have contributed to the contrast effect (Manis and Moore, 1978). As discussed earlier, Chien et al (2010) suggest

that contrast effects are more likely when the plausible ranges of the context and target stimuli are perceived to be non-overlapping. In both experiments where participants made a forecast for both scenarios we examined the extent to which their prediction intervals overlapped. In experiment 1, where a moderate context scenario was presented, the prediction intervals were non-overlapping 55% of the time; when an extreme context scenario was presented they were non-overlapping 81% of the time. In experiment 2, 69% of the intervals did not overlap. Hence a perception of non-overlapping scales may also have contributed to the observed contrast effects when the context scenario was extreme.

One possibility is that the contrast effects observed in Group 3 (who saw two scenarios and produced forecasts for both) in both experiments merely reflected a superficial tendency of the participants to set their two point forecasts widely apart to indicate that they had read both scenarios. Indeed, the failure to observe a contrast effect for group 2 (who saw two scenarios but only produced forecasts for one of them) in Experiment 2 would be consistent with this. Give that this group saw both a context and target scenario, but only produced a forecast for the latter, they would have had no need to distinguish their forecast from one based on the opposing scenario. However, the results of experiment 1 provide evidence that this superficial tendency did not apply. In this experiment, those who saw two scenarios made the forecast for the target scenario first so, at this point, there was no forecast for the context scenario from which they might try to diverge –though the participants would have been aware that a forecast for the context scenario would subsequently be required (see Appendix 1). Also, as shown in Table 1, moderate context scenarios led to a smaller contrast effect then more extreme context scenarios, as predicted by the literature (e.g. Sherif et al., 1958). If the participants were merely arbitrarily setting their two forecasts widely apart, then there was no reason to expect this effect. In addition, recall that a contrast effect was not observed in the participants' rating of the scenarios in experiment 2. Had the participants simply been signalling that they had read the scenarios, it seems likely that the groups who saw both scenarios would also have produced more divergent ratings, as well as more divergent point forecasts.

This raises the question of why the act of merely rating the context scenario in the task carried out by group 2 in experiment 2, who saw two scenarios but only made a forecasts for one of them, did not lead to a contrast effect like that observed in group 3, who produced forecasts for both scenarios. Recall that the ratings were made on scale from -5 (extremely pessimistic) to +5 (extremely optimistic). It is possible that, when the participants rated the pessimistic and optimistic scenarios, they regarded the negative and positive parts of the scale as two separate scales -a pessimism scale and an optimism scale. For example they would be unlikely to think in terms, such as: 'this best-case scenario is six points more optimistic than the alternative worst-case scenario'. As such, the rating element of the task, unlike the forecasting element, would not lead to a direct comparison of the scenarios. The focus would be on each scenario in turn. This would be unlikely to engender dissimilarity-orientated, or comparative, processing. For each country, we also calculated the correlation of the participants' rating for the target scenario with their point forecast. The

mean correlations (averaged across the scenarios) ranged from only 0.16 to 0.24 for the three groups, suggesting that the ratings had little relationship with the forecasting element of the task. This further supports the idea that the rating and forecasting aspects of the task drew on different modes of thinking.

5.2 Prediction intervals

The higher confidence, represented by the narrower intervals around the point forecasts of the people in group 3 in both experiments who were presented with two scenarios is consistent with earlier findings that confidence in forecasts can increase as more, and even irrelevant, information, becomes available to the forecaster (e.g. Oskamp, 1965; Hall et al., 2007). Normatively, the opposing forecast was irrelevant when determining the range between the worst and best case forecast for the target scenario. Again, it is possible that the narrower intervals produced by participants who saw two scenarios in both experiments merely reflected a desire to keep the prediction intervals for the two scenarios apart by avoiding an overlap. However, as we pointed out earlier, in experiment 1 the prediction interval for the target scenario was elicited before that of the context scenario so, at this stage, there was no alternative prediction interval with which overlap needed to be avoided. In experiment 2, despite having access to information about two scenarios, the mean width of intervals produced by group 2 in experiment 2 was not significantly narrower than that of group 1, who only saw one scenario. This may be because the act of merely rating the context scenario did not make the information in this scenario as salient as it was for members of group 3, who produced a forecast for this scenario, but further work would be needed to establish this.

Other studies have found that there were unwarranted increases in confidence in predictions when people were supplied with scenarios (Schnaars and Topol, 1987; Kuhn and Sniezek, 1996; Önkal, Sayım and Gönül, 2013). However, it is important to distinguish between a prediction interval constructed with each limit in turn being based on an optimistic and a pessimistic scenario respectively, as in these studies and a prediction interval based on the assumption that a single scenario will prevail –as in the current study. While these earlier studies have reported a tendency for forecasters to be overconfident -in that their intervals are overly narrow -the findings of a contrast effect in the current study may actually offer a means of counteracting overconfidence in that the presentation of an opposing scenario can push the target best and worst case forecasts to more extreme values. For example, eliciting a best case forecast under an optimistic scenario (with a pessimistic scenario also presented) and then a worst case forecast under the pessimistic scenario (with the optimistic scenario also still presented) should lead to the production of wider interval than one where the opposing scenarios were not present.

Of course, further research would be needed to establish how far overconfidence would be mitigated and whether there is a danger that the resulting interval would be so wide that it manifested under confidence. However, the extent to which the procedure can lead to wide intervals is evidenced by those obtained from participants in Group 3 in experiment 1. Recall that participants in this group produced

forecasts for the two scenarios they saw. When, in each case, the context and target scenarios were extreme the mean difference between their best case forecast, based on the optimistic scenario, and worst case forecast, based on the pessimistic scenario, was 5.4 times the standard deviation of noise (recall this was 20 units) in the demand series. Of course, the potential uncertainty about the future demand level would be determined not only by noise, but also by events like those described in the scenarios. However, for comparison, it is worth noting that a statistically-based 99% prediction interval, given this level of noise, would have a width of 5.2 standard deviations.

5.3 Limitations

Our study has a number of limitations. The scenarios were entirely qualitative. For example, in experiment 1, they did not contain quantitative data such as potential market size, demographic statistics, market research figures and likely prices relative to those of competitors. Nor did they contain historic statistics, which showed the past impact of factors like competition or regulatory changes, or probabilities that given scenarios would ensue. In these respects they are typical of scenarios often used in strategic planning (van der Heijden, 1996) where high levels of uncertainty preclude the estimation of specific values. For example, consider these phrases used in scenarios produced by Statoil, a Norwegian oil and gas company (Ringland, 1998 p.253). 'The gradual resolution of budget and trade deficits result in higher levels of growth worldwide', 'oil prices rebound', 'commodity and energy prices plunge', 'material and energy intensity declines dramatically'. There are no specifications of what the future levels of growth, prices or intensity will be. The emphasis is on direction and magnitude is expressed in semantic terms. In addition, probabilities are absent because the role of the scenarios is not to specify the relative likelihood of particular futures, but to establish bounds on the range of possible futures that may play out.

Nevertheless, the participants may have been justified in asking, for example: how do I translate a statement like 'Improvements to the model's touchscreen and battery life are successfully implemented and praised in the media' into a quantitative forecast and many of their freeform comments in the exit questionnaire to experiment 1 hinted at such questions. While experience in companies (Fildes, Goodwin, Lawrence and Nikolopoulos, 2009) suggests that forecasts are often based on qualitative information like our scenarios, there may be an underlying incompatibility between the qualitative nature of much scenario planning and the quantitative requirements of demand forecasts. Despite this we believe that our results are interesting in that our main aim was to detect the direction of any context effect, rather than its magnitude. It is even possible that the observed effect may have been more marked if the scenarios had contained more specific quantitative information as the literature review indicated that less vague stimuli are more likely to result in contrast effects.

Secondly, while we asked individuals to make forecasts in both our experiments, often forecasts are agreed in social situations, such as forecast review meetings in companies (Fildes et al., 2009). In this case,

different managers may present alternative scenarios –often based on their specialism - and the ensuing discussion will aim to take these into account. In these situations the resolved forecast may also be dependent on factors such as the force of personality and status of individual managers and the salience their scenarios have within the discussion.

A third potential limitation is that the time series in our experiments covered only a few months and were purely random. However, again this reflects many practical contexts where product life cycles are short (e.g., Yelland, 2010; Hankammer and Steiner, 2015) – particularly for the type of products used in experiment 1. In such short series, characteristics such as trends and seasonal patterns may be difficult to discern anyway. Recall also that the role of the series was simply to provide guidance on typical levels and variation of the variables to be forecast so that the focus could be on the implications of the scenarios.

A final potential limitation of this study is that the results are based on the participation of students in a laboratory experiment, rather than practising managers making forecasts at their place of work. However, the laboratory provided the benefits of a controlled environment so that other issues, such as company or national politics or having vested interests in the outcome, could be excluded. Moreover, several studies have suggested that management students can act as reliable proxies for the behaviour of practising managers (Remus, 1986, Bolton, Ockenfels and Thonemann, 2012; Trottier and Gordon, 2016).

6.0 Conclusions

The main finding of this study is that, when individuals made assessments of the implications of scenarios, their forecasts were subject to context effects in the form of contrast effects. This conflicts with the normative way in which scenarios should be treated –that is, they should be regarded as belonging to separate independent worlds so that one scenario does not have any implications for another.

Our findings suggest a number of possible pathways for future research. We found that presenting two opposing scenarios implicitly increased participants' confidence in their forecasts based on the target scenario, in that the prediction intervals they produced were narrower. However, our study was not designed to investigate the effect of this on the calibration of the resulting prediction intervals. These considerations are important because recent research has suggested that well calibrated prediction intervals can improve decision making (Ramos, Van Andel and Pappenberger, 2013; Savelli and Joslyn, 2013).

Secondly, while our results may have implications for scenario planning methods that generate just two scenarios such as the extreme world method (e.g., Goodwin and Wright, 2014), it would be interesting to examine the extent to which context effects are likely to occur when other scenario planning methods, which yield more than two scenarios, are applied. For example, the intuitive logics method normally generates four scenarios with each scenario representing factors that are associated with extremes of impact or uncertainty (Wright, Bradfield and Cairns, 2013).

Thirdly, the participants in our experiments had no role in the formulation of the scenarios. It would be interesting to investigate whether the same effects would be observed in situations where forecasters are also involved in the development and writing of scenarios.

Fourth, because scenarios often represent sequences of events that unfold as a target future date is approached, their relationship to psychological work on how people perceive sequences may also be worth investigating (e.g. see Oskarsson, Van Boven, McClelland and Hastie, 2009). For example, if people believe in the gambler's fallacy, they may perceive that a scenario that portrays a series of unmitigated negative events is implausible and attach less weight to its implications, or alternatively, they may judge that a turn in fortunes is overdue and make their forecasts accordingly.

Given that the focuses of our study have been on medium-term demand and stock market forecasting, further research would be needed to establish the extent to which the findings apply in other situations. For example, it might even be extended to examine the effects of presenting of positive and negative scenarios in advertising designed to encourage people to make better health-related decisions (Krishen and Bui, 2015). If context effects are found to be common and potentially damaging when scenarios are being used as part of decision making processes, then it will be necessary to adopt methods that are likely to mitigate these effects while benefiting from the richness provided by alternative scenarios when judgments are being elicited. Efforts to support and enhance scenario technologies in decision making will benefit from examining the comparative impact of alternative scenario generation techniques on context effects.

Acknowledgements

The authors would like to thank three anonymous referees and the Editor for their helpful and insightful comments on earlier versions of this paper.

Appendix 1. An example of an instruction sheet used in Experiment 1 for one product

Group 3

Name: ……

University: ……

TechGeek Company

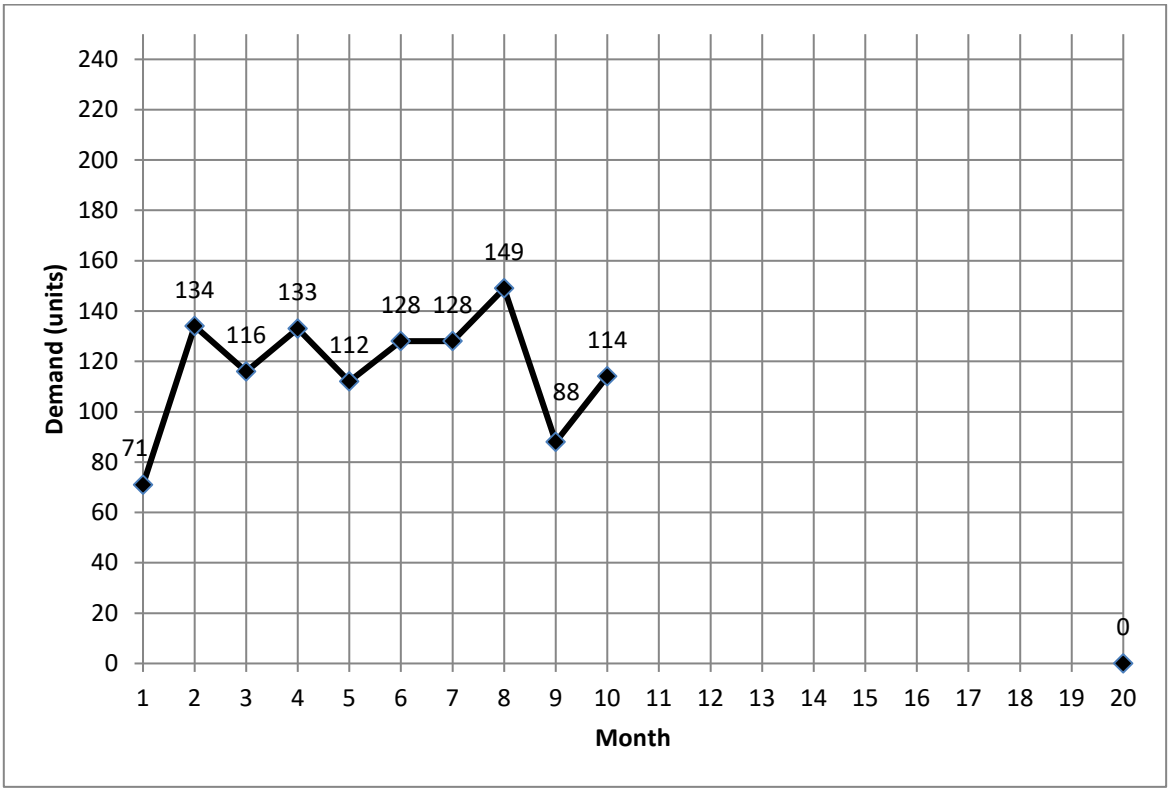
TechGeek Company manufactures high technology products (such as drones, smart phones, wearable fitness devices etc.) and offers technology related services such as high-speed internet and movie/series streaming. Competition in this industry is very harsh in general. One of the major costs is related to research and development (R&D) for new products and innovations. As new technologies are being constantly introduced to the market, speed is essential for remaining competitive in today's digital era. Moreover, as new products and services are introduced into the market very quickly, existing ones have a very short span of life. If not sold, such 'aged' products increase the overall costs drastically. Finally, there are other potential costs to be minimised for technological products: firstly, the cost of over-production and the associated storage costs; and secondly, the cost of lost sales for to running out of stock as a result of under-production. Combined together, these factors amplify the importance of forecasting the demand correctly (with minimal error) for *TechGeek Company*'s various products and services.

In order to improve the accuracy of its forecasts, *TechGeek* employs a combination approach that involves using the scenario method. Ten-month-ahead demand forecasts are generated at the beginning of each month (e.g., in the beginning of January demand forecasts of the coming November are constructed). This process works as follows:

1. "Worst case" and/or "best case" scenarios for each product and service's demand in the coming 10-month period are developed by a team of experts in the company using all the information available. The basic rule in preparing and using these scenarios is that every scenario is equally plausible - these scenarios are constructed to depict 'best' and 'worst' plausible futures that need to be considered and prepared for regarding each product and service's potential demand for the next period.

2. The scenarios (generated by the expert team) are given to executives in charge of making forecasts. These people produce various forecasts (a point forecast, a best-case forecast and a worst-case forecast) in light of the scenario(s) and the demand history provided for each product and service. Your task in this study replicates this process

Product A: Tablet computer **Group 3**



Worst case scenario

We start to receive many complaints that the built-in applications (apps) do not work properly with upgrades to our operating system. In addition, we receive reports that many touch screens are causing severe problems after a long period of use. A competitor launches a rival product which is slightly cheaper than Product A and offers significant advantages over it, including a more stylish look and superior screen display. The rival product receives much publicity and many positive reviews in the media. Market research reports indicate that potential new purchasers strongly favour the rival product.

Best case scenario

Product A has an expanding and loyal user base. It is improved with some significant and attractive modifications and more built-in applications (apps) and these receive enthusiastic reviews in industry magazines and in the wider media. In particular, the product continues to be regarded as the best of its type for its reliability and touch screen display. No models directly compete with this product in its market. Market research reports indicate that the product's popularity is very high and that it is likely to continue to attract a significant number of new purchasers.

Product A: Tablet computer **Group 3**

Assume that the **worst case** scenario applies to market conditions in month 20

Please provide a *point forecast* for demand in month 20 : …….

Please provide a *best-case forecast (highest value predicted)* for month 20 : ……

Please provide a *worst-case forecast (lowest value predicted)* for month 20 :…….

Assume that the **best case** scenario applies to market conditions in month 20

Please provide a *point forecast* for demand in month 20 : …….

Please provide a *best-case forecast (highest value predicted)* for month 20 : ……

Please provide a *worst-case forecast (lowest value predicted)* for month 20 :…….

Appendix 2. An example of an instruction sheet used in Experiment 2 for one country

Group 3

Name: ……

University: ……

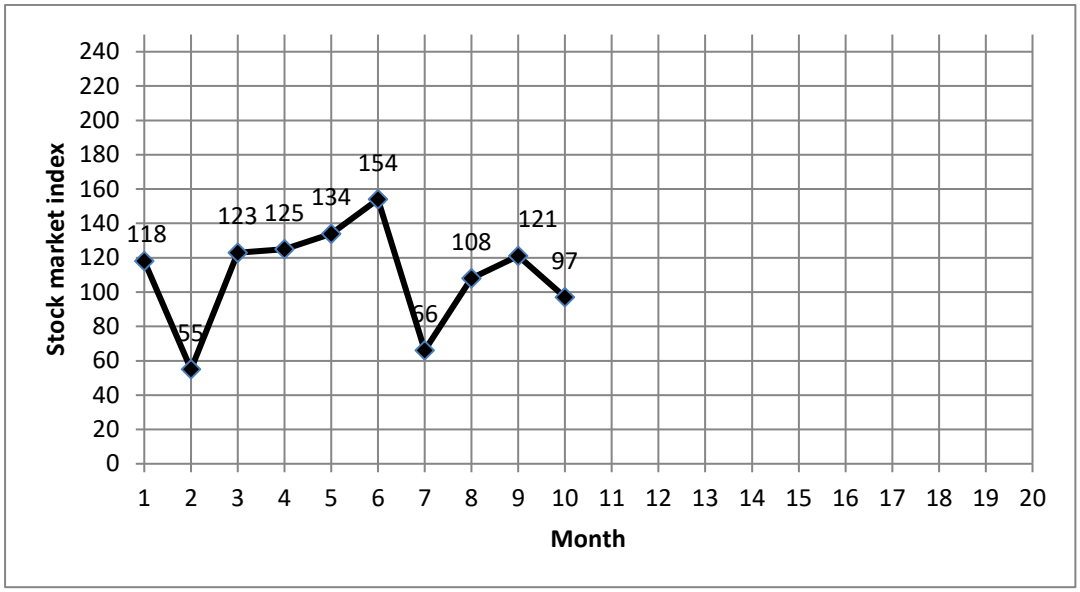
EconGlobal Company

EconGlobal Company is an international consulting company that generates stock market index forecasts for various countries to provide guidance to investors all over the world. Given the overall volatility of macroeconomical conditions worldwide, in order to improve the accuracy of its forecasts, *EconGlobal* employs a combination approach that involves using the scenario method. Ten-month-ahead stock market index forecasts are generated at the beginning of each month (e.g., in the beginning of January stock market index forecasts of the coming November are constructed). This process works as follows:

1. "Worst case" and/or "best case" scenarios for each country's overall economic outlook in the coming 10-month period are developed by a team of experts in the company using all the information available. The basic rule in preparing and using these scenarios is that every scenario is equally plausible - these scenarios are constructed to depict 'best' and 'worst' plausible futures that need to be considered and prepared for regarding each country's potential economy for the coming periods.

2. The scenarios (generated by the expert team) are given to executives in charge of making forecasts. These people produce various forecasts (a point forecast, a best-case forecast and a worst-case forecast) in light of the scenario(s) and the stock market index history provided for each country. Your task in this study replicates this process.

Country: Tegoland **Group 3**



BEST CASE scenario for Month 20

The new government's much anticipated 'grow-the-economy' budget is already proving to be effective with retail sales soaring according to the latest Bloomberg monthly survey of retail purchasing managers. The positive economic outlook and declining oil prices have also boosted consumer confidence and household spending, encouraged by the US-led boom in the world economy. Growth in the Tegoland economy has been particularly responsive to these buoyant international conditions and some economists think that this quarter may see the highest growth rates for the last quarter of a century.

Rate this scenario for its degree of pessimism or optimism (circle the appropriate number)

Extremely Pessimistic	-5	-4	-3	-2	-1	0	1	2	3	4	5	Extremely Optimistic

WORST CASE scenario for Month 20

Retail sales in Tegoland declined in most of the last 6 months as a bleaker economic outlook damped consumer demand, according to the latest Bloomberg retail purchasing managers index monthly survey. Record oil prices have pushed up fuel costs, crimping household spending. Retail sales will tumble this quarter amid worsening economic conditions. The global outlook is darkening by the day, and Tegoland's low potential growth rate makes it particularly vulnerable to these dire international economic conditions. It looks almost certain that this quarter will be recorded as one of the deepest recessions in the last quarter of a century

Rate this scenario for its degree of pessimism or optimism (circle the appropriate number)

Extremely Pessimistic	-5	-4	-3	-2	-1	0	1	2	3	4	5	Extremely Optimistic

References

Ben-David, I., Graham, J. R., & Harvey, C. R. (2013). Managerial miscalibration. *Quarterly Journal of Economics*, 128, 1547-1584.

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

Bolton, G. E., Ockenfels, A., & Thonemann, U. W. (2012). Managers and students as newsvendors. *Management Science*, *58*(12), 2225-2233.

Chernev, A., & Blair, S. (2015). Doing well by doing good: the benevolent halo of corporate social responsibility. *Journal of Consumer Research*, *41*(6), 1412-1425.

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.

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

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.

Förster, J., Liberman, N., & Kuschel, S. (2008). The effect of global versus local processing styles on assimilation versus contrast in social judgment. *Journal of Personality and Social Psychology*, *94*(4), 579.

Goodwin, P. (2014). Getting real about uncertainty. *Foresight: The International Journal of Applied Forecasting*, (33), 4-7.

Goodwin, P. & Wright, G. (2014). *Decision Analysis for management judgment*, (5th ed.) Chichester UK: Wiley.

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.

Hankammer, S., & Steiner, F. (2015). Leveraging the sustainability potential of mass customization through product service systems in the consumer electronics industry. *Procedia CIRP*, *30*, 504-509.

Harvey, N., & Reimers, S. (2013). Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 589.

Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of experimental social psychology*, *19*(4), 323-340.

Kelman, M., Rottenstreich, Y., & Tversky, A. (1996). Context-dependence in legal decision making. *The Journal of Legal Studies*, *25*(2), 287-318.

Krishen, A. S., & Bui, M. (2015). Fear advertisements: Influencing consumers to make better health decisions. International Journal of Advertising, *34*(3), 533-548.

Kuhn, K.M. & Sniezek, J.A. (1996). Confidence and uncertainty in judgmental forecasting: differential effects of scenario presentation. *Journal of Behavioral Decision Making*, 9, 231–247.

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

Louie, K., Khaw, M. W., & Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, *110*(15), 6139-6144.

Manis. M., & Moore, J. C. (1978) Summarizing controversial messages: Retroactive effects dueto subsequent information. *Social Psychology*, *41*, 62-68.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250.

O'Donnell, E., & Schultz Jr, J. J. (2005). The halo effect in business risk audits: Can strategic risk assessment bias auditor judgment about accounting details? *The Accounting Review*, *80*(3), 921-939.

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.

Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological bulletin*, *135*(2), 262.

Pizzi, G., & Scarpi, D. (2013). When out-of-stock products do backfire: Managing disclosure time and justification wording. *Journal of Retailing*, *89*(3), 352-359.

Plous, S. (1993). *The psychology of judgment and decision making*. New York: McGraw Hill Inc.

Remus, W. (1986). Graduate students as surrogates for managers in experiments on business decision making. *Journal of Business Research*, *14*(1), 19-25.

Ringland, G. (1998). *Scenario Planning. Managing for the Future.* Chichester: Wiley.

Ramos, M. H., Van Andel, S. J., & Pappenberger, F. (2013). Do probabilistic forecasts lead to better decisions? *Hydrology and Earth System Sciences*, *17*, 2219-2232.

Savelli, S., & Joslyn, S. (2013). The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied Cognitive Psychology*, *27*(4), 527-541.

Scarpi, D. and Pizzi, G. (2013). The impact of phantom decoys on choices and perceptions. *Journal of Behavioural Decision Making*, *26*(5), 451-461.

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

Schwarz, N., & Bless, H. (1992a). Constructing reality and its alternatives: Assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 217–245). Hillsdale, NJ: Erlbaum.

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

Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, *29*(3), 281-295.

Soltani, A., De Martino, B., & Camerer, C. (2012). A range-normalization model of context-dependent choice: a new model and evidence. *PLoS Comput Biol*, *8*(7), e1002607.

Stapel, D. A., & Winkielman, P. (1998). Assimilation and contrast as a function of context-target similarity, distinctness, and dimensional relevance. *Personality and Social Psychology Bulletin*, *24*(6), 634-646.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology, 53*, 1-26.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social psychology*, *73*(3), 437.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*(1), 25-29.

Trottier, K., & Gordon, I. M. (2016). Students as surrogates for managers: Evidence from a replicated experiment. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*.

Trueblood, J. S., & Pettibone, J. C. (2017). The phantom decoy effect in perceptual decision making. *Journal of Behavioral Decision Making*, *30*(2), 157-167.

van der Heijden, K. (1996). *Scenarios. The Art of Strategic Conversation.* Chichester: Wiley.

Vlaev, I., Chater, N., Stewart, N., & Brown, G. D. (2011). Does the brain calculate value? *Trends in Cognitive Sciences*, *15*(11), 546-554.

Welsh, M. B., & Begg, S. H. (2018). More-or-less elicitation (MOLE): reducing bias in range estimation and forecasting. *EURO Journal on Decision Processes*, *6*(1-2), 171-212.

Wright, G., Bradfield, R., & Cairns, G. (2013). Does the intuitive logics method–and its recent enhancements–produce "effective" scenarios? *Technological Forecasting and Social Change*, *80*(4), 631-642.

Yelland, P. M. (2010). Bayesian forecasting of parts demand. *International Journal of Forecasting*, *26*(2), 374-396.

	Group	Target scenario	Context scenario	Mean Point forecast	SD
	1	Extreme pessimism	None	98.3	21.0
	2	Extreme pessimism	Moderate optimism	88.7	18.7
a)	3	Extreme pessimism	Extreme optimism	79.4	18.8
		$F_{2,111}$= 9.1		p <0.0005	η^2 = 14.1%

	Group	Target scenario	Context scenario	Mean Point forecast	SD
	1	Moderate pessimism	None	107.3	17.5
	2	Moderate pessimism	Moderate optimism	103.5	13.6
b)	3	Moderate pessimism	Extreme optimism	97.8	15.6
		$F_{2,111}$= 3.6		p=0.03	η^2 = 6.1%

	Group	Target scenario	Context scenario	Mean Point forecast	SD
	1	Moderate optimism	None	137.9	18.3
c)	2	Moderate optimism	Moderate pessimism	148.3	19.1
	3	Moderate optimism	Extreme pessimism	158.3	20.7
		$F_{2,111}$= 11.0		p <0.0005	η^2 = 16.5%

	Group	Target scenario	Context scenario	Mean Point forecast	SD
	1	Extreme optimism	None	135.0	21.6
	2	Extreme optimism	Moderate pessimism	149.7	17.5
d)	3	Extreme optimism	Extreme pessimism	152.1	17.3
		$F_{2,111}$= 19.5		p <0.0005	η^2 = 14.6%

Table 1 ANOVAs comparing mean point forecasts in Experiment 1

	Group	Target scenario	Context scenario	Mean width	SD
	1	Extreme pessimism	None	55.2	19.0
	2	Extreme pessimism	Moderate optimism	41.4	18.7
a)	3	Extreme pessimism	Extreme optimism	37.9	16.5
		$F_{2,111}$= 10.1		p <0.0005	η^2 = 15.4%

	Group	Target scenario	Context scenario	Mean width	SD
	1	Moderate pessimism	None	44.5	17.3
	2	Moderate pessimism	Moderate optimism	39.9	12.6
b)	3	Moderate pessimism	Extreme optimism	32.9	17.1
		$F_{2,111}$= 35.3		p=0.006	η^2 = 8.7%

	Group	Target scenario	Context scenario	Mean width	SD
	1	Moderate optimism	None	57.6	18.4
c)	2	Moderate optimism	Moderate pessimism	44.1	23.2
	3	Moderate optimism	Extreme pessimism	38.5	19.8
		$F_{2,111}$= 9.2		p <0.0005	η^2 = 14.2%

	Group	Target scenario	Context scenario	Mean width	SD
	1	Extreme optimism	None	48.2	19.1
	2	Extreme optimism	Moderate pessimism	40.5	17.5
d)	3	Extreme optimism	Extreme pessimism	38.1	18.2
		$F_{2,111}$= 3.3		p=0.041	η^2 = 5.6%

Table 2 Mean width of prediction intervals in Experiment 1

Group	Target scenario	Context scenario	Mean point forecast	SD
1	Extreme pessimism	None	93.4	15.0
2	Extreme pessimism	Extreme optimism	99.8	18.1
3	Extreme pessimism	Extreme optimism	78.9	15.0
	$F_{2,87} = 13.4$		$p<0.0005$	$\eta^2 = 23.5\%$

Group	Target scenario	Context scenario	Mean point forecast	SD
1	Extreme optimism	None	111.9	16.2
2	Extreme optimism	Extreme pessimism	102.0	18.2
3	Extreme optimism	Extreme pessimism	118.3	16.1
	$F_{2,87} = 7.1$		$p=0.001$	$\eta^2 = 14.0\%$

Table 3 ANOVAs comparing mean point forecasts in Experiment 2

Group	Target scenario	Context scenario	Mean interval width	SD
1	Extreme pessimism	None	59.0	25.4
2	Extreme pessimism	Extreme optimism	72.1	32.0
3	Extreme pessimism	Extreme optimism	36.8	20.1
	$F_{2,87} = 13.64$		$p < 0.0005$	$\eta^2 = 22.1\%$

Group	Target scenario	Context scenario	Mean interval width	SD
1	Extreme optimism	None	74.2	28.1
2	Extreme optimism	Extreme pessimism	74.0	26.0
3	Extreme optimism	Extreme pessimism	44.5	22.5
	$F_{2,87} = 13.3$		$p < 0.0005$	$\eta^2 = 21.7\%$

Table 4 Mean width of prediction intervals in Experiment 2

Group	Target scenario	Context scenario	Mean rating of target scenario	Mean rating of context scenario
1	Extreme pessimism	None	-2.7	n/a
2	Extreme pessimism	Extreme optimism	-1.8	1.9
3	Extreme pessimism	Extreme optimism	-2.8	2.7
			p =0.02	p =0.02

Group	Target scenario	Context scenario	Mean rating of target scenario	Mean rating of context scenario
1	Extreme optimism	None	2.8	n/a
2	Extreme optimism	Extreme pessimism	2.3	-2.6
3	Extreme optimism	Extreme pessimism	2.4	-2.3
			p =0.18	p =0.48

Ratings of scenarios from -5 (extreme pessimism) to +5 (extreme optimism)

p -values are from ANOVAs comparing means for each group

Table 5 Mean ratings of scenarios in Experiment 2