In press: *International Journal of Forecasting*

Using Scenarios to Forecast Outcomes of a Refugee Crisis

Lars Wicke[1], Mandeep K. Dhami[2*], Dilek Önkal[3], and Ian K. Belton[4]

[1]London School of Economics and Political Science, UK

[2]Middlesex University, London, UK

[3]University of Northumbria, UK

[4]University of Strathclyde, UK

*Send correspondence to:
Mandeep K. Dhami
Department of Psychology
Middlesex University
The Burroughs, Hendon
London, UK
NW4 4BT
E-mail: m.dhami@mdx.ac.uk

Acknowledgements

This research was supported by funding to Dhami by HM Government. We would like to thank David Wasely for his research assistance.

Abstract

The Syrian civil war has led to millions of Syrians fleeing the country, and has resulted in a humanitarian crisis. By considering how such socio-political events may unfold, scenarios can lead to informed forecasts that can be used for decision-making. We examined the relationship between scenarios and forecasts in the context of the Syrian refugee crisis. Forty Turkish students trained to use a brainstorming technique generated scenarios that might follow within six months of the Turkish government banning Syrian refugees from entering the country. Participants generated from 3-6 scenarios. Over half were rated as 'high' quality in terms of completeness, relevance/pertinence, plausibility, coherence, and transparency (order effects). Scenario quality was unaffected by scenario quantity. Even though no forecasts were requested, participants' first scenarios contained from 0-17 forecasts. Mean forecast accuracy was 45% and this was unaffected by forecast quantity. Therefore, brainstorming can offer a simple and quick way of generating scenarios and forecasts that can potentially help decision-makers tackle humanitarian crises.

Keywords: Scenario generation, forecasting, brainstorming, refugees, humanitarian crisis

Introduction

Scenario generation and forecasting are central to decision-making in many domains. In the domain of intelligence analysis, decision-making to tackle security threats and related socio-political problems is based on an analysis of how a current situation may evolve over a given timeframe or what might prevail at some point in the future given a set of factors. In recent years, the effectiveness of the intelligence community's ability to generate accurate scenarios and forecasts has been called into question. For example, the Arab Spring which began in December 2010 in Tunisia was a surprise to many Western Governments (e.g., Goodwin, 2011; Morrell, 2015). In addition, few expected the initial protests in Syria in January 2011 to evolve into an ongoing civil war. And still, few were prepared for the refugee crisis that resulted from this civil war. For several years now, the world has witnessed the plight of the fleeing Syrian population caught up in conflict and has arguably been slow to react to the unfolding humanitarian crisis. Official reactions to this crisis, such as countries accepting Syrian refugees on mass, have themselves led to an array of unanticipated outcomes such as the rise in pro-national fundamentalism in host countries.

The ability of analysts to construct plausible scenarios that can help identify how the future may unfold is of value to decision-makers responding at both strategic and tactical levels to such complex and dynamic socio-political events. By considering causal processes that aim to uncover how an event may unfold over time, it may be argued that scenario generation can result in fairly well informed or reasoned forecasts. Indeed, recent research on improving the accuracy of analytic forecasts has demonstrated that training in scenario generation confers some benefits (e.g., Chang, Chen, Mellers, & Tetlock, 2016; Mellers et al., 2014). Similarly, some of the extant literature on scenario generation has discussed the utility of scenarios for forecasting (e.g., Goodwin & Wright, 2010; Kuhn & Sniezek, 1996; Önkal, Sayim, & Gonul, 2013; Schnaars & Topol, 1987; Wright & Goodwin, 2009).

Despite this acknowledgement of the intrinsic link between scenario generation and forecasting, few researchers have explored both issues simultaneously (for exceptions see Cui et al., 2015 and Zhou, Tesfastion, & Liu, 2009). Indeed, scenario generation and forecasting are sometimes considered as quite distinct tasks within the intelligence community. For instance, the structured analytic techniques (SATs) that analysts are trained and encouraged to use differ for both of these tasks (for a review see Dhami, Belton, & Careless, 2016). Dhami et al. (2016) noted that SATs which have the primary function of aiding scenario generation do not have the primary function of aiding prediction/forecasting, and vice, versa.

In the present paper, we examine scenario generation and forecasting in relation to the Syrian refugee crisis. We empirically assess the quality of the scenarios generated and consider the accuracy of any forecasts made during the process of scenario generation. Forecasting the outcomes of events such as the Syrian refugee crisis is crucial to planning for, and responding to these dynamic events (Comfort, Sungu, Johnson & Dunn, 2001). Indeed, host nations need to arrange (temporary) shelter, food, water, sanitation and health services, as well as organize resettlement. They also need to anticipate the impact that refugees may have, for example, on their economy and population. Scenario generation can be useful because it can describe best or worst possible outcomes, that are more or less surprising, in the near-, medium- or long-term, across a variety of domains (e.g., economic, health). Thus, scenarios can lead to a range of forecasts that can help decision-makers to adopt a proactive rather than reactive response to what can quickly become a humanitarian crisis. Before we describe the present method and findings, we provide a brief review of the relevant literature on scenario generation and its linkages to forecasting.

Scenario Generation

Scenario generation is considered necessary when one needs to know what can happen, or how a state of the future can be achieved (Borjeson, Hojer, Drebord, Kevall, &

Finnveden, 2006). There are both quantitative and qualitative methods for generating scenarios and each type of method has its advantages and disadvantages (Amer, Daim, & Jetter, 2013). The intelligence community typically applies qualitative methods for scenario generation. Such methods are characterised by an intuitive logic approach (Bradfield Bradfield, Wright, Burt, Cairns, & Van Der Heijden, 2005; Bunn & Salo, 1993, Ducot & Lubben, 1980; Foster, 1993; Huss & Honton, 1987).

In a review of 75 SATs used by the intelligence community, Dhami et al. (2016) found that there were several SATs with the primary function of helping analysts generate scenarios. These included alternative futures analysis, cone of plausibility, morphological analysis, multiple scenarios generation, outside-in-thinking, simple scenarios, and (individual/unstructured, group/virtual structured) brainstorming. The SATs differed in terms of the expertise, time and resources required to implement them.

Perhaps the most commonly used technique is individual brainstorming (e.g., UK Government, 2013; US Government, 2009), which also forms part of some other techniques such as alternative futures analysis (Heuer & Pherson, 2014). This technique requires analysts to write down as many thoughts as possible that come to mind in response to a specific question. It relies on the analyst's ability to think creatively or laterally, and to think freely, without self-censorship. Brainstorming is not only relatively simple, but it also requires minimal training, time and resources. It can be employed by individuals as well as by groups, and can be used repeatedly during the analytic process.

Although SATs such as brainstorming often form part of the core skill set taught in analytic training programs (e.g., Marrin, 2008) because they are meant to encourage transparent and critical thinking, they have not been sufficiently empirically examined (e.g., Dhami, Mandel, Mellers, & Tetlock, 2015). In a review of 50 studies conducted from 1958 to 1988 on brainstorming, Isaksen (1998) concluded that much of the research focused on comparing individual and group brainstorming rather than focusing on the main purpose of this technique i.e., to generate useful scenarios.

Most studies on brainstorming use the quantity of generated ideas as a measure of the effectiveness of this technique. Isaksen's (1998) review found that 32 out of 50 studies used quantity as an evaluation measure, with only 15 using quality and 10 using originality as measures (see also Al-Samarraie & Hurmuzan, 2018). Paulus, Kohn, and Arditti (2011) found that asking participants to focus on generating a large number of ideas lead to better quality ideas. Here, quality was defined as the novelty or uniqueness of an idea and its utility or potential for positive impact (both criteria had to be rated as 3 or above on a 5-point rating scale in order for an idea to be classed as high quality).

The quality of scenarios generated using brainstorming, or any other technique, can be judged based on a variety of criteria. In a review of 15 studies, Amer et al., (2013; see also Al-Samarraie & Hurmuzan, 2018) found the following criteria were used to judge scenarios, listed here in order of popularity: plausibility, consistency/coherence, relevance/pertinence, creativity/novelty, completeness, transparency and importance. In addition, some have judged scenarios based on their redundancy (e.g., Furnham & Yazdanpanahi, 1995; Litchfield, Fan & Brown, 2011; see also Al-Samarraie & Hurmuzan, 2018). Finally, Nowack, Endrikat, and Guenther (2011) suggest that scenarios should be credible (in terms of internal validity and reliability), transferable (in terms of external validity), and legitimate (from the perspective of the users).

It could be argued that brainstorming, with its emphasis on lateral and creative thinking, may sometimes yield scenarios that fall short on some of the aforementioned criteria, namely, practicality, relevance/pertinence and plausibility. On the other hand, brainstorming may yield scenarios that are high on creativity/novelty.

Scenarios in Aid of Forecasting

Scenarios can provide information that is useful for forecasting because they encourage thinking about causal processes that may unfold over time (Önkal et al., 2013) and for planning because they can identify basic trends and uncertainties (e.g., Shoemaker, 1995). Suleimenova, Bell and Groen (2017) for example, used simulated scenarios to forecast refugee movements, and others have used scenarios to investigate peace-keeping missions in refugee camps (Johnson, Lampe & Seichter, 2009). Asking decision-makers to construct their own scenarios provides an effective channel into future-focused thinking (Wright & Cairns, 2011; Wright & Goodwin, 1999). Scenarios may also be considered to have higher credibility than dry statistical projections which may explain why they have a strong impact on judgments about the future (Schnaars & Topol, 1987; Taylor & Thompson, 1982).

In recent research, Mellers et al. (2014) trained participants to generate scenarios before forecasting geopolitical events. "Scenario training taught forecasters to generate new futures, actively entertain more possibilities, use decision trees, and avoid biases such as over-predicting change, creating incoherent scenarios, or assigning probabilities that exceed 1 to mutually exclusive and exhaustive outcomes." (p. 1107). This training was found to significantly increase forecast accuracy (as measured by Brier scores).

In related work, Chang et al. (2016) provided scenario training to individuals asked to make geopolitical forecasts. The "Scenario-training was a four-step process: (1) developing coherent and logical probabilities under the probability sum rule; (2) exploring and challenging assumptions; (3) identifying the key causal drivers; (4) considering the best and worst case scenarios and developing a sensible 95% confidence interval of possible outcomes; and (5) avoid over-correction biases." (p. 513). The training was devised to encourage "trainees to think critically about assumptions, potential futures, and causal mechanisms that could be at play on a given forecasting question." (p. 513). The training was found to have a significant positive effect on forecast accuracy (measured by Brier scores).

However, some researchers have pointed out that scenario generation using techniques such as brainstorming may make some outcomes more imaginable and so can distort forecasts by making less plausible outcomes appear more credible and more likely (Chang & Tetlock, 2016; Chang, Berdini, Mandel, & Tetlock, 2018). While focusing on a single scenario can cause frame blindness (Wright & Rowe, 2011), introduction of a second scenario can lead to a ‘scenario-dilution effect’ whereby the forecasts based on the original scenario could be shifted excessively in the opposite direction implied by the new scenario (Önkal, Goodwin, & Gonul, 2017).

In sum, although it is widely acknowledged that the task of scenario generation can aid the task of forecasting, there is also some concern about the accuracy of the forecasts that might be provided in these scenarios. However, we are not aware of any empirical studies investigating the direct relationship between scenario generation and socio-political forecasting. The present paper aims to fill this gap by not only measuring the nature and number of forecasts that may be made as a result of the task of scenario generation using the brainstorming technique, but by also measuring the accuracy of those forecasts. This can help us to determine the effectiveness of the brainstorming technique for scenario generation and explore its viability as an approach to forecasting, particularly in the context of socio-political events and humanitarian crises.

The Present Study

The main aims of the present study were to: (1) measure the quantity of brainstormed analytic scenarios generated in response to a specific analytic task; (2) assess the quality of these scenarios in relation to a specific set of a priori criteria; (3) examine the relationship between scenario quantity and quality; (4) measure the number of forecasts provided in these scenarios; (5) describe the nature of these forecasts in terms of a specific set of a priori criteria; (6) assess the accuracy of these forecasts, if possible[1]; and (7) measure the relationship between forecast accuracy, number and nature.

Method

Participants

Participants were 40 final year undergraduate students who elected to study the 'State, Society and Citizenship' course at a Turkish University. Twenty-three of the sample were female. The mean age of the sample was 22.10 (*SD* = .98).

Individual Brainstorming Technique Training

Participants were trained to use the individual brainstorming technique described by Heuer and Pherson (2014). This comprises four steps as follows: Choose a comfortable place to think and minimize distractions; clearly define the focal question; write down as many thoughts as possible in response to that question, however odd they may sound, do not self-censor; arrange responses into appropriate categories in order of importance to the issue of interest. A mind map may be used to arrange and develop ideas. The worked example that formed the training is provided in the Appendix.

Analytic Task

The analytic task on the topic of Syrian refugees was selected because it was considered pertinent to participants, and it was believed that they would have some general knowledge of the topic given its relevance to the country in which they lived. In addition, the research team believed that there was some chance that the event mentioned in the task might actually occur at some point after data collection, thus enabling the accuracy of any forecasts provided to be measured.

[1] At the time of designing the study, we were uncertain if the future state of the world described in the analytic task for which participants were asked to generate scenarios would actually prevail, and whether the scenarios generated by our participants would contain any forecasts.

The task instructed participants as follows: “We want you to imagine that the Turkish government decides to ban all Syrian refugees from entering Turkey, effective immediately. Using the scenario generation technique you have been taught, we would like you to think about the scenarios (as many as you can) that might occur as a consequence of this policy decision within the next 6 months. Please use your own knowledge, and, if you want, you can also read some of the background information on the Syrian refugee crisis that follows.”

The information sheet summarised how the civil war in Syria and the refugee crises started, what each involved party is fighting for, how Syrian people are affected and how Turkey and other countries responded to the crises. This information was drawn from a variety of online sources. (Readers should contact the second author for a copy of this information sheet).

Participants were instructed to write each scenario on a separate page and to number their scenarios in the order they generated them. Note that participants were not specifically asked to provide forecasts in their scenarios.

Measures and Data Coding

We coded both the quality of the brainstormed scenarios and the nature and accuracy of any forecasts contained in these scenarios. The variables to be coded were generated by the research team following a review of the extant literature on scenario generation and forecasting. Only variables that could be reliably (consistently) coded were used, and this was assessed by testing re-testing the coder(s). The variables and coding are described below.

Criteria for assessing scenario quality. The quality of the scenarios was assessed using five criteria taken from the past literature. *Completeness* of the scenario refers to the detail of the drivers, outcomes, assumptions, and provision of relevant background information. The scenario *context* variable (otherwise known as ‘relevance/pertinence’) refers to the current social, economic, legal, political context, as well as past history. Scenario *plausibility* refers to the link between the drivers and outcomes in the scenario, as well as assumptions. *Coherence* of the scenario refers to the logical flow of the argument presented in the scenario. The *order effects* variable (otherwise known as 'transparency') refers to the identification of 2nd or 3rd order effects in the scenario. Each variable had three levels as follows:

- Completeness was coded as 'high' if the scenario contained a description of the drivers as well as the outcomes, showed some consideration of assumptions, and included relevant background information; 'medium' if it contained a description of the drivers and outcomes, but no assumptions or background information; and 'low' if it only contained a description of outcomes or drivers and/or assumptions or background information.
- *Context* was coded as 'high' if the scenario referred to three or more contexts (e.g., political, economic, security); 'medium' if it referred to two contexts; and 'low' if it referred to only one or no context.
- *Plausibility* was coded as 'high' if the scenario made explicit/clear links between specific drivers and specific outcomes, as well as links between specific drivers, outcomes and the assumptions underlying them; 'medium' if it made links between drivers and outcomes and assumptions that were not totally clear/explicit or specific; and 'low' if it made no links between drivers and outcomes.
- *Coherence* was coded as 'high' if the ideas contained in the scenario were all consistent with each other and not contradictory; 'medium' if the scenario contained an argument with one inconsistency/contradiction; and 'low' if the scenario contained more than one inconsistency/contradiction.
- *Order effects* was coded as 'high' if the scenario specified the potential effects of all the outcomes; 'medium' if the scenario specified the potential effects of only some of

the outcomes; and 'low' if the scenario did not mention the potential effects of any outcomes.

Variables describing nature of forecasts. The nature of the forecasts was described in terms of five variables (i.e., domain, outcome direction, detail, precision, and uncertainty). These had the following levels:

- *Domain* was coded as '1' if the forecast referred to economic conditions (including tourism, employment and education); '2' if it concerned politics (including immigration); and '3' if it related to the social domain (including crime/security and health/medical).
- *Outcome direction* was coded as '1' if the forecasted outcome was positive; '2' if it was negative; and '3' if it was neutral. Examples of neutral outcomes are: "There would be more economic regulation in Turkey" (Pp 13-prediction 1) and "It might affect unemployment rate" (Pp 31-prediction 2). The former makes no indication of whether the change would be positive or negative, and the latter makes no indication of whether the change refers to an increase or decrease.
- *Detail* was coded as '1' if only the outcome was mentioned in the forecast; '2' if the outcome plus one direct cause was mentioned; and '3' if the outcome plus a chain of causes were mentioned.
- *Precision* was coded as '1' if the forecast was general and '2' if the forecast was specific. An example of a general forecast is "There are other countries which help Syria" (Pp 4-prediction 3). This is general because there is no mention of the countries. An example of a specific forecast is "*Putin government might directly intervene and worse situation*".
- *Uncertainty* was coded as '1' if the forecast included phrases indicating uncertainty (e.g., "might") and '2' if no such language was used. The uncertainty language was

taken from previous literature investigating verbal expressions of probability and uncertainty in intelligence analysis (see Kesselman, 2008).

Measurement of forecast accuracy. Data collection was in early December 2015, and the decision by Turkey was made a few weeks later. Thus, the accuracy of forecasts was measured six months after (i.e., from January to June 2016). This was done by referring primarily to an a priori list of online sources that were deemed to be reliable by the research team (e.g., BBC, Guardian newspaper, World Bank etc). A forecast was coded as accurate if there was evidence from a reliable source suggesting that the forecasted outcome occurred within the six-month timeframe, and a forecast was coded as inaccurate if there was no evidence found from a reliable source to suggest that the forecasted outcome had occurred within the six-month timeframe. Note, this represents a conservative measure of forecast accuracy for at least two reasons: one, because the forecasted outcome might have occurred outside the six-month timeframe, and two, because the research team may not have found evidence from a reliable source to suggest that the forecasted outcome had occurred within the six-month timeframe even though it did.

Procedure

Participation was voluntary, without any remuneration, and included informed consent and debriefing procedures. Data was collected individually. Participants were seated on separate tables in a typical auditorium, equipped with pencil and paper. Participants first received a short tutorial on how to use the brainstorming technique. This was provided by the third author. Participants all indicated that they understood the technique and had no further questions. They were then presented with the analytic task described above and had the opportunity to ask questions before starting. Participants were free to leave when they felt they had finished writing their scenarios (however many that was). The maximum duration for participation was approximately 110 minutes.

Analysis and Findings

The 40 participants wrote a total of 135 scenarios. Participants' first scenarios alone (*n* = 40) contained 176 forecasts in total. The data analyses and findings are described below in order of the study aims listed earlier.

Quantity and Quality of Brainstormed Scenarios

We first determined the number of scenarios generated by each participant. Then, we calculated the proportion of scenarios across participants that were assigned each level of the five variables measuring quality (i.e., completeness, context, plausibility, coherence, and order effects) described in the Method section above. Finally, we measured the association between the number of scenarios generated across participants and the median quality of the scenarios.

Scenario quantity (N = 40 participants). Participants generated from three to six scenarios (*M* = 3.30, *SD* = .72).

Scenario quality (N = 135 scenarios). Figure 1 illustrates the proportion of scenarios that were coded as 'high', 'medium' and 'low' quality on each of the five criteria for assessing quality. As can be seen, over half of the scenarios were coded as 'high' on each of the criteria. Specifically, across participants, 56.0% percent of scenarios were coded as 'high' on completeness. They contained a description of the drivers as well as the outcomes, showed some consideration of assumptions, and included relevant background information. Sixty percent of scenarios were coded as 'high' on the context criteria, meaning that they made reference to three or more contexts (e.g., political, economic, security). In terms of plausibility, 51.9% of scenarios were coded as 'high' because they made explicit/clear links between specific drivers and specific outcomes, as well as links between specific drivers, outcomes and the assumptions underlying them. Virtually all (98.5%) of the scenarios were coded as 'high' on coherence because the ideas contained in them were all consistent with each other and not contradictory. Finally, 61.9% of scenarios were coded as 'high' in terms of identification of order effects because the potential effects of all the outcomes were specified.

FIGURE 1 ABOUT HERE

Only a tiny minority of scenarios were coded as 'low' on each of the criteria. Specifically, across participants, 3.0% of scenarios were coded as 'low' on completeness because they only contained a description of outcomes or drivers and/or assumptions or background information; 3% referred to only one context; 4.4% of scenarios made no links between drivers and outcomes; and 3.0% of scenarios did not mention the potential effects of any outcomes. None of the scenarios contained more than one inconsistent/contradictory argument (i.e., coded as 'low' on coherence).

Relationship between scenario quantity and quality. In order to measure the association between the number of scenarios generated and the quality of the scenarios, we first computed the median of the five criteria measuring quality (i.e., completeness, context, plausibility, coherence, and order effects) across the scenarios generated by each participant. Then, we analysed each criteria separately using Spearman's rho. There was no statistically significant association between the number of scenarios generated by participants and the median rating of each measure of quality (i.e., context, plausibility, and order effects; rho could not be computed for completeness and coherence because of the lack of variability in these two variables) across participants, all ps > .05. Thus, scenario quality was not contingent on scenario quantity.

Number, Nature and Accuracy of Forecasts

In order to explore how the task of scenario generation may relate to forecasting, we further analysed the contents of the first scenario generated by each participant. The first scenario represents a direct implementation of the brainstorming technique, whereas later scenarios may be anchored on earlier ones (and so not a direct result of using the technique).

The first scenario was, therefore, used to measure the number of forecasts made, the nature of the forecasts, and the accuracy of the forecasts.

We first determined the number of forecasts provided by each participant in their first scenario. We then calculated the proportion of forecasts across participants (i.e., N = 176 forecasts) that were assigned each level of the five variables describing the nature of the forecasts (i.e., domain, outcome direction, detail, precision, and uncertainty) listed in the Method section above. Next, we determined the accuracy of the forecasts using the procedure described in the Method section. Following this, across participants, we measured the association between average forecast accuracy and the number of forecasts provided. We also measured the relationship between forecast accuracy and the nature of the forecasts.

Forecast number (N = 40 participants, n = 40 scenarios). Participants provided from 0 to 17 forecasts (M = 4.40, SD = 3.72) in their first scenario.

Forecast nature (N = 176 forecasts). Across participants' first scenarios, 44.3% of forecasts were in the political/immigration domain, 37.5% were in the social domain (including crime/security and health), and 18.2% of forecasts were in the economic domain (including tourism and education). Nearly half (48.3%) of the forecasts referred to an outcome occurring in a negative direction, 27.8% were positive and 23.9% were neutral. Sixty-one percent of forecasts mentioned the outcome only, and the remainder (39.4%) mentioned the outcome plus one cause. None mentioned the outcome plus a chain of causes. The majority (78.2%) of forecasts were general rather than specific (21.8%). Words conveying a degree of (un)certainty appeared in 72.6% of forecasts. More specifically, the language used included the following terms: will, would, can, might, could, maybe, and if.

Forecast accuracy. Forty-four percent of the 176 forecasts were found to be accurate (and 55.7% inaccurate). Taking into account the number of forecasts generated by each participant, across participants, the mean accuracy rate was 44.67% (SD = 32.43%).

Relationship between forecast number, nature and accuracy. Across participants, there was no significant correlation between the number of forecasts provided by each participant and the average accuracy of the forecasts, $r = .010$, $p = .955$. Thus, forecast accuracy was not contingent on forecast number.

In order to explore the association between forecast accuracy and the nature of the forecasts, crosstabs were first generated to show the percentage of forecasts that were accurate or inaccurate at each level of the five variables describing the nature of the forecasts (i.e., domain, outcome direction, detail, precision, and uncertainty). The findings are presented in Figure 2.

FIGURE 2 ABOUT HERE

Next, binary logistic regression analysis was performed where variables were entered in two blocks. Following a constant only model, the first block controlled for participant and forecast, while the second block measured the (predictive) association between the five variables describing the nature of the forecasts and forecast accuracy. The final model was statistically significant, χ^2 (9) = 31.96, $p < .001$. Prediction success rose from 56.6% to 64.7% and Nagelkerke's R^2 was .23. The Wald criterion demonstrated that three of the variables i.e., forecast domain, outcome direction and detail contributed significantly to the predictive utility of the model, all *p*s < .05 (see Figure 2).

Discussion

The movement of fleeing peoples around the world has become one of the new century's main challenges. According to the UNHCR (2019), there were 25.4 million refugees in 2017, and 6.5 million of these were from Syria. Turkey was the top refugee-hosting country. Forecasting the impact of a refugee crisis, such as the one caused by the Syrian civil war, and understanding the outcomes that may emerge from such dynamic and unstable socio-political events is crucial for host countries such as Turkey. Future-focused thinking tasks such as constructing plausible scenarios and forecasting potential outcomes can help decision-makers respond in a proactive and adaptive way. For instance, they can more accurately estimate the amount of humanitarian aid and support that is necessary to accommodate refugees. Decision-makers can also try to mitigate any negative effects a large influx of refugees may have on host nations. Although generating scenarios and constructing forecasts are intrinsically related, researchers and practitioners have often considered them separately when exploring ways to understand and improve analytic performance and their impact.

In the present paper, using an analytic task that is representative of one faced by strategic intelligence analysts, we examined both scenario generation and forecasting in relation to the Syrian refugee crisis. Specifically, we empirically assessed the quality of the scenarios generated, and the accuracy of any forecasts that were made during the process of scenario generation. We did this by asking Turkish students to perform one analytic task. The present findings need to be replicated using other samples and other analytic tasks to explore the generalizability of the results. The present study revealed three sets of findings and these are summarized next.

The first set of findings refers to the task of scenario generation. Our participants were not asked to focus on (or aim for) quantity or quality. Despite this, we found that participants were able to fairly easily and rapidly generate from three to six scenarios. Amer et al. (2013) suggest that three to five scenarios are sufficient, and Pillkahn (2008) argues that four scenarios provide the best cost-benefit ratio (where cost refers to drafting and evaluating them and benefit refers to quality). In addition, we found that over half of the scenarios were rated as ‘high’ quality in terms of their completeness, context (relevance/pertinence), plausibility, coherence, and order effects (transparency), and very few were of ‘low’ quality. Past studies have found that quantity (rather than quality or joint quantity-quality) instructions are beneficial for scenario generation (e.g., Paulus et al., 2011). Thus, this sort of direction could conceivably have increased scenario quality even further in the present study. Perhaps most importantly, we also found that scenario quality was not affected by the number of scenarios generated. The high degree of motivation of participants in the present study may partly explain why there was little quality loss associated with generating scenarios, and we may expect equally motivated analysts to perform similarly.

The second set of findings pertains to the task of forecasting. Although our participants were not instructed to provide forecasts, most provided several forecasts in just one scenario alone. The nature of these forecasts varied in terms of their domain, directionality, complexity of causal processes, level of generality, and degree of uncertainty. Forty-four percent of the forecasts were revealed to be accurate within the six-month time frame examined. Accuracy was significantly associated with domain, outcome direction and detail. A slightly longer timeframe may have yielded increased accuracy. Indeed, in a study of real numerical forecasts made by strategic analysts, Mandel and Barnes (2014) found a high a degree of accuracy (measured as calibration and discrimination) for forecasts given for 0-6 months and 6-12 months ahead. Again, we found that forecast accuracy was not affected by the number of forecasts generated. To-date, there is mixed evidence on the relationship between forecast accuracy (variously measured) and the number of forecasts generated (Chang et al., 2016), and our finding adds to this literature.

The fact that participants in the present study were lay people with no specific expertise in the analytic task, no training in forecasting, and who were given only fairly simple brainstorming training, begs the question: How can expert intelligence forecasting performance be improved? Mellers and her colleagues (2015a, b) suggest that performance can be improved by a combination of selecting the right sort of person and providing the right opportunities. Experts would need specific cognitive abilities (defined in terms of fluid intelligence, numeracy and cognitive reflection), political knowledge, open-mindedness (assessed via actively open-minded thinking, need for closure, and hedgehog-fox orientation). They would also need deliberation time (i.e., time spent before making a forecast), frequent belief updating (measured by the number of forecasts made over time for each item), motivation and commitment (defined in terms of deliberate practice and focused efforts towards skill cultivation, belief updating and information gathering), and training (which could be in the form of scenario-training and/or training directed towards probabilistic reasoning and cognitive debiasing; see also Chang, et al, 2016; Mellers et al., 2014).

The final set of findings to emerge from the present study refers to the relationship between scenario generation and forecasting. Although participants were not instructed to provide forecasts, 35 (out of 40) did so. This lends some support for the argument that scenario generation is a precursor to forecasting. Scenarios can aid forecasting partly because they provide a natural language for thinking about events in a causal way. Our participants provided fairly elaborate scenarios, nearly two thirds of which were rated as ‘high’ in terms of the identification of order effects (i.e., specified the potential effects of the outcomes mentioned). The fact that most scenarios contained a causal chain of events means that they can be practically useful for decision-makers.

Before we discuss the potential implications of the present findings, it is worth pointing out that our participants provided responses in English (which is not their native language, but is the medium of communication at their university). This may have adversely affected the length and quality of their scenarios and forecasts. Thus, the above findings likely represent a conservative measure of the quantity and quality of brainstormed analytic scenarios and the number and nature of forecasts they may contain. A promising direction for further work would entail examining the relationship between language (and verbal) ability and performance on scenario generation and forecasting tasks.

Potential Implications and Directions for Future Research

Anticipating events such as the Arab Spring can be particularly challenging due to the wide array of evolving factors involved (Bayat, 2013; but see Asongu & Nwachukwu, 2016). Foreseeing the consequences of such a ‘surprising’ event in terms of, for example, the ensuing Syrian civil war and the resulting Syrian refugee crisis is even more difficult. However, scenario generation using the brainstorming technique may provide a relatively simple and quick way to generate good quality scenarios and appears to be a reasonable approach to making forecasts about a humanitarian crisis such as the one studied here.

Scenarios describe complex events in terms of specific drivers, and this can provide the level of detail necessary for making precise forecasts. Scenarios can also pertain to a variety of domains (e.g., economic, health), and so forecasts will be relevant to a range of events. In addition, scenarios may present best or worst possible states of the world, which means that forecasts may vary in their level of optimism or pessimism. Scenarios may also be about events that are more or less likely, and so forecasts can differ in the degree of uncertainty attached to them. Finally, scenarios can cover near-, medium- or long-term events, and so forecasts may pertain to a range of time horizons.

We have shown that brainstorming can be a useful alternative to expensive and time consuming techniques, programmes and simulations. Efficient and effective techniques are particularly valuable during a humanitarian crisis such as a refugee crisis, where time is of the essence and critical decisions need to be made swiftly. Some other qualitative scenario generation techniques currently in use in the intelligence community such as cone of plausibility and simple scenarios (Heuer & Pherson, 2014) require analysts to generate a baseline scenario, along with a best-case or optimistic scenario as well as the worst-case or pessimistic scenario, and even a ‘wildcard’ scenario. It is unknown whether these techniques result in good quality scenarios, and what implications they have for the accuracy of any forecasts that may be contained in them. Such scenarios provide information about the 'tail' or extreme boundaries of what could occur, and so may be low on plausibility. On the other hand, such scenarios could be used to (appropriately) narrow the range of forecasts made. Recently, it has been pointed out that some aspects of the Arab Spring were considered in a 'worst case' scenario developed by the US intelligence community, but the forewarning was ignored by decision-makers (Arcos & Palacios, 2018). Thus, even if the extreme scenarios generated using some techniques can result in accurate forecasting, their seeming lack of plausibility may result in decision-makers ignoring the information provided.

Collaborative analysis is also common in the intelligence analysis domain (Dhami & Careless, 2015), and this may include group brainstorming. Some past studies have shown that group brainstorming is less productive in terms of the quantity and quality of idea generation than individual brainstorming (e.g., Camacho & Paulus, 1995; Furnham, 2000; Mullen, Johnson, & Salas, 1991; Paulus & Dzindolet, 1993). Others have shown that group-to-individual sequencing of brainstorming sessions can be beneficial as they allow individuals to exploit ideas generated by the group (Paulus, Korede, Dickson, Carmeli, & Cohen-Meitar, 2015). These studies, however, have not considered the accuracy of the forecasts that may be contained in these scenarios. Mellers et al. (2015b) found that working in 'cognitively enriched team environments' (where members are given training on strategies for building effective teams, explaining their thinking and forecasts to others, offering constructive feedback and information sharing) is associated with improved forecast accuracy (see also Mellers et al., 2014). Further work on the effect of group brainstorming on the number and accuracy of forecasts contained in the scenarios generated would be useful.

In sum, given the prevailing uncertainties in complex socio-political events, constructing plausible scenarios can help to identify how the future may unfold. This can be tremendously useful to those who must make decisions that respond to these events in strategic and tactical ways that are both proactive and adaptive. The time sensitive nature of humanitarian crises that ensue from socio-political events means that decisions need to be made in a timely manner. Therefore, techniques that are both effective and efficient are warranted, and future research ought to focus further on scenario-forecast synergies.

References

Al-Samarraie, H., & Hurmuzan, S. (2018). A review of brainstorming techniques in higher education. *Thinking Skills and Creativity*, *27*, 78-91.

Alcamo, J. (Ed.) (2008). *Environmental futures: The practice of environmental scenario analysis.* UK: Elsevier Science.

Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures*, *46*, 23-40.

Arcos, R., & Palacios, J. (2018). The impact of intelligence on decision-making: The EU and the Arab Spring. *Intelligence and National Security, 33*(5), 737-754.

Asongu, S. A., & Nwachukwu, J. C. Revolution empirics: predicting the Arab Spring. *Empirical Economics, 51*(2), 439-482.

Bayat, A. (2013). The Arab spring and its surprises. *Development and Change, 44*(3), 587-601.

Börjeson, L., Höjer, M., Dreborg, K. H., Ekvall, T., & Finnveden, G. (2006). Scenario types and techniques: towards a user's guide. *Futures*, *38*(7), 723-739.

Bradfield, R., Wright, G., Burt, G., Cairns, G., & Van Der Heijden, K. (2005). The origins and evolution of scenario techniques in long range business planning. *Futures*, *37*(8), 795-812.

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

Camacho, L. M., & Paulus, P. B. (1995). The role of social anxiousness in group brainstorming. *Journal of personality and social psychology*, *68*(6), 1071.

Chang, W., Berdini, E., Mandel, D. R., & Tetlock, P. E. (2018). Restructuring structured analytic techniques in intelligence. *Intelligence and National Security, 33*(3), 337-356.

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.

Chang, W., & Tetlock, P. (2016). Rethinking the training of intelligence analysts. *Intelligence and National Security, 31*(6), 903-920.

Comfort, Louise K., Yesim Sungu, David Johnson, & Mark Dunn (2001). Complex systems in crisis: Anticipation and resilience in dynamic environments. *Journal of Contingencies and Crisis Management, 9*(3), 144-158.

Cui, M., Ke, D., Sun, Y., Gan, D., Zhang, J., & Hodge, B. M. (2015). Wind power ramp event forecasting using a stochastic scenario generation method. *IEEE Transactions on Sustainable Energy*, *6*(2), 422-433.

Dhami, M. K., Belton, I. K., & Careless, K. (2016, August). Critical review of analytic techniques. In *Intelligence and Security Informatics Conference (EISIC), 2016 European (pp.*152-155) IEEE.

Dhami, M. K., & Careless, K. E. (2015). Intelligence analysis: Does collaborative analysis outperform the individual analyst? *The Journal of Intelligence Analysis*, *22*(3), 43-58.

Dhami, M. K., Mandel, D. R., Mellers, B., & Tetlock, P. (2015). Improving intelligence analysis with decision science. *Perspectives on Psychological Science, 10*(6), 753-757.

Ducot, G., & Lubben, G. J. (1980). A typology for scenarios. *Futures*, *12*(1), 51-57.

Foster, M. J. (1993). Scenario planning for small business. *Long Range Planning 26*, 123-129.

Furnham, A. (2000). The brainstorming myth. *Business strategy review*, *11*(4), 21-28.

Furnham, A., & Yazdanpanahi, T. (1995). Personality differences and group versus individual brainstorming. *Personality and Individual Differences*, *19*(1), 73-80.

Goodwin, J. (2011). Why were we surprised (again) by the Arab spring? *Swiss Political Science Review, 17*(4), 452-456

Goodwin, P., & Wright, G. (2010). The limits of forecasting methods in anticipating rare events. *Technological Forecasting and Social Change*, *77*(3), 355-368.

Heuer, R. J., & Pherson, R. H. (2014). *Structured analytic techniques for intelligence analysis*. Washington DC: CQ Press.

Huss, W. R., & Honton, E. J. (1987). Scenario planning—what style should you use?. *Long range planning*, *20*(4), 21-29.

Isaksen, S. G. (1998). *A review of brainstorming research: Six critical issues for inquiry*. Buffalo, NY: Creative Research Unit, Creative Problem Solving Group-Buffalo.

Johnson, R. T., Lampe, T. A., & Seichter, S. (2009). Calibration of an agent-based simulation model depicting a refugee camp scenario. In *Simulation Conference (WSC), Proceedings of the 2009 Winter* IEEE (pp. 1778-1786).

Kesselman, R. F. (2008). *Verbal probability expressions in national intelligence estimates: A comprehensive analysis of trends from the fifties to post 9/11*. Unpublished Masters dissertation, Mercyhurst College.

Kuhn, K. M., & Sniezek, J. A. (1996). Confidence and uncertainty in judgmental forecasting: differential effects of scenario presentation. *Journal of Behavioral Decision Making, 9*(4), 231-247.

Litchfield, R. C., Fan, J., & Brown, V. R. (2011). Directing idea generation using brainstorming with specific novelty goals. *Motivation and Emotion*, *35*(2), 135-143.

Mandel, D. R., & Barnes, A. (2014). Accuracy of forecasts in strategic intelligence. *Proceedings of the National Academy of Sciences of the United States of America, 111*(30), 10984-10989.

Marrin, S. (2008). Training and educating U.S. intelligence analysts. *International Journal of Intelligence and Counterintelligence*, *22*(1), 131-146.

Mellers, B. A., Stone, E., Atanasov, P., Roghbaugh, N., Metz, S. E., Ungar, L.,… & Tetlock, P. E. (2015a). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of Experimental Psychology: Applied*, *21*, 1-14.

Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M, & Ungar, L. (2015b). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science*, *10*, 267-281.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Guracy, B., Fincher, K., … & Murray, T. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, *25*(5), 1106-1115.

Morrell, M. (2015). *The Great War of Our Time: The CIA's Fight against Terrorism--From al Qa'ida to ISIS*. New York: Twelve, Hachette Book Group.

Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and applied social psychology*, *12*(1), 3-23.

Nowack, M., Endrikat, J., & Guenther, E. (2011). Review of Delphi-based scenario studies: quality and design considerations. *Technological Forecasting and Social Change*, *78*(9), 1603-1615.

Önkal, D., Goodwin, P., & Gonul, S. (2017). *Effects of scenario tone on judgmental forecast adjustments*. INFORMS Annual Meeting 2017, Houston.

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

Paulus, P. B., & Dzindolet, M. T. (1993). Social influence processes in group brainstorming. *Journal of Personality and Social Psychology, 64*(4), 575.

Paulus, P. B., Kohn, N. W., & Arditti, L. E. (2011). Effects of quantity and quality instructions on brainstorming. *The Journal of Creative Behavior*, *45*(1), 38-46.

Paulus, P. B., Korde, R. M., Dickson, J. J., Carmeli, A., & Cohen-Meitar, R. (2015). Asynchronous brainstorming in an industrial setting: Exploratory studies. *Human factors*, *57*(6), 1076-1094.

Pillkahn, U. (2008). *Using trends and scenarios as tools for strategy development: shaping the future of your enterprise*. Chichester, UK: John Wiley & Sons.

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. (1993). Multiple scenario development: Its conceptual and behavioral foundation. *Strategic management journal*, *14*(3), 193-213.

Schoemaker, P. J. (1995). Scenario planning: a tool for strategic thinking. *Sloan management review*, *36*(2), 25.

Suleimenova, D., Bell, D., & Groen, D. (2017). A generalized simulation development approach for predicting refugee destinations. *Scientific reports*, *7*(1), 13377.

Taylor, S. E., & Thompson, S. C. (1982). Stalking the elusive" vividness" effect. *Psychological review*, *89*(2), 155.

UNHCR (2019). *Figures at a glance.* Retrieved from https://www.unhcr.org/uk/figures-at-a-glance.html

UK MOD (2013). *Quick wins for busy analysts*. London, UK: Defence Intelligence, UK.

US Government (2009). *A tradecraft primer: Structured analytic techniques for improving intelligence analysis.* Retrieved from https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/Tradecraft%20Primer-apr09.pdf

Wright, G. & Cairns, G. (2011). *Scenario thinking: practical approaches to the future.* Hampshire, UK: Palgrave Macmillan.

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.

Wright, G. & Rowe, G. (2011). Group-based judgmental forecasting: an integration of extant knowledge and the development of priorities for a new research agenda. *International Journal of Forecasting*, 27, 1-1.

Zhou, Q., Tesfatsion, L., & Liu, C. C. (2009, March). Scenario generation for price forecasting in restructured wholesale power markets. In *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES* (pp. 1-8). IEEE.

Figure 1. **Scenario quality**

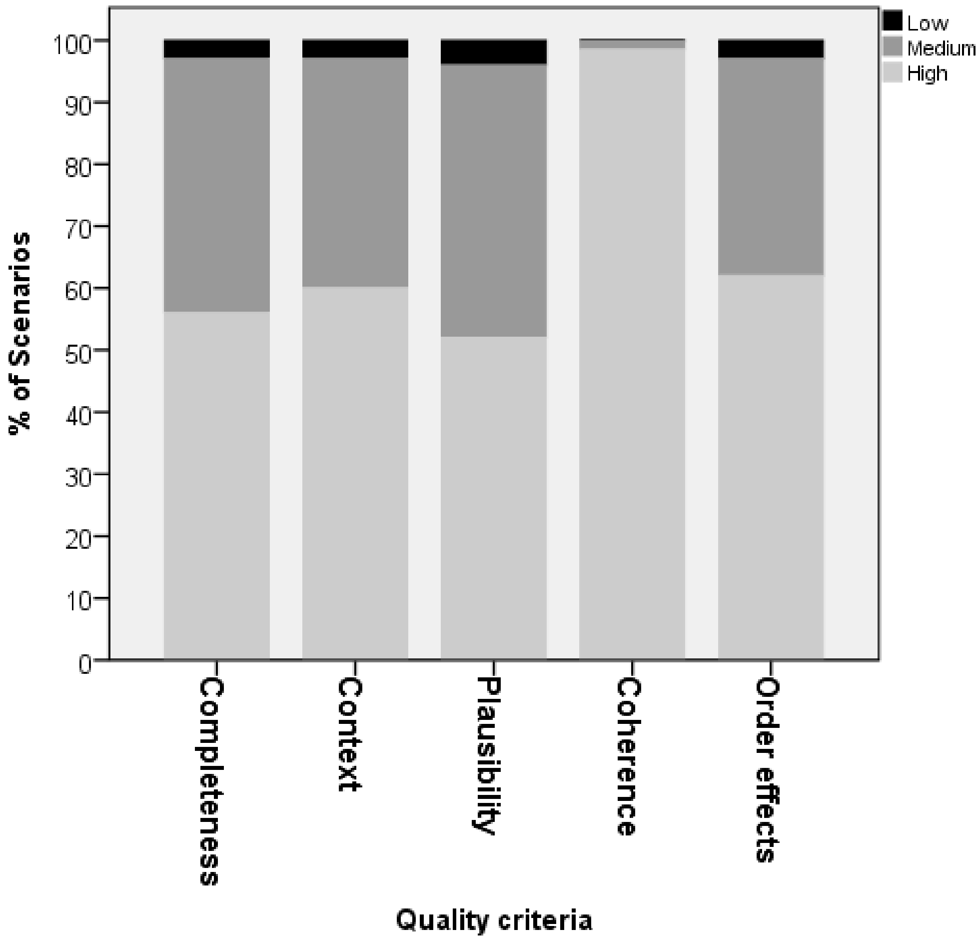
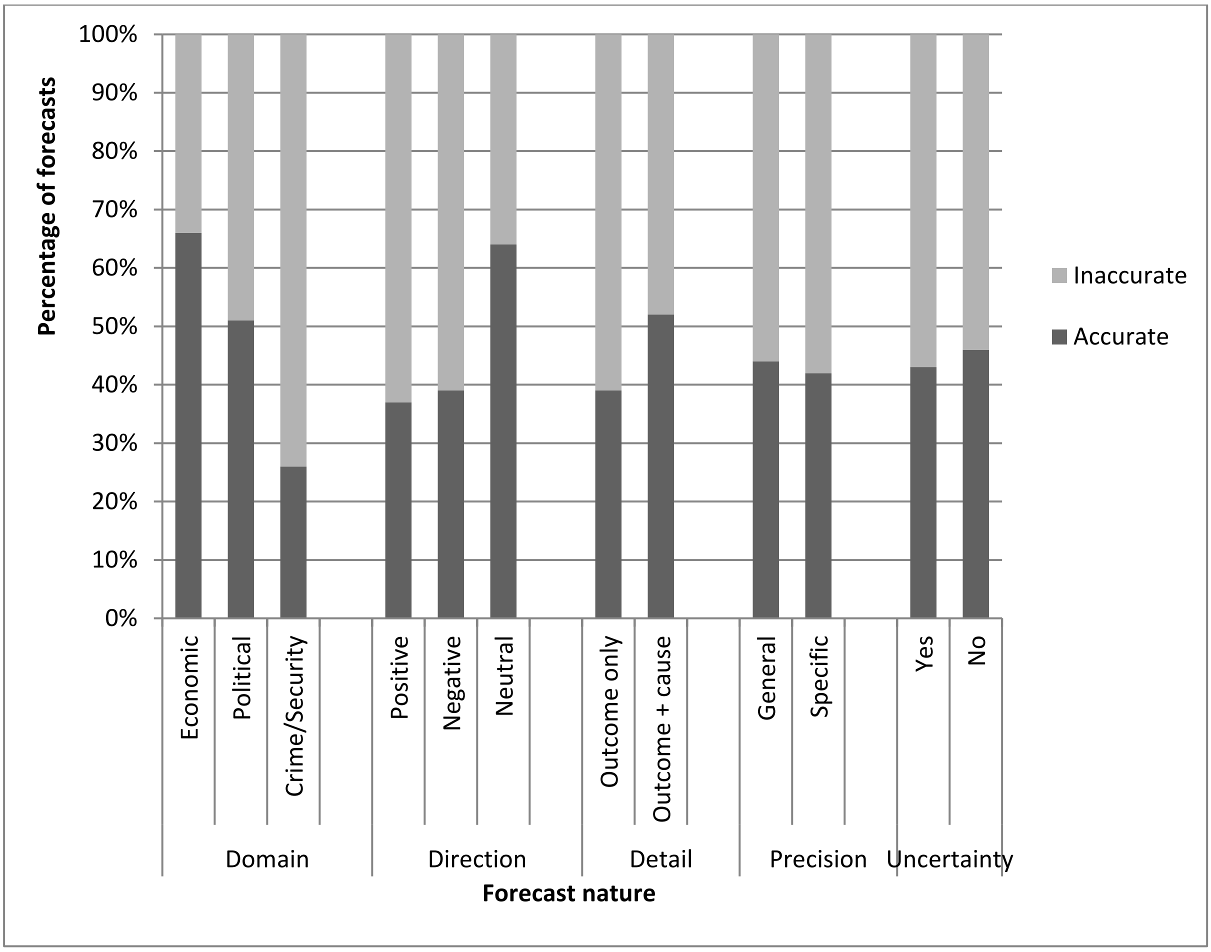


Figure 2. **Percentage of accurate and inaccurate forecasts by nature of forecasts**



Appendix

INDIVIDUAL BRAINSTORMING – TRAINING MATERIALS

Worked Example

Step 1: Choose a comfortable place to think and minimize distractions.

Step 2: Clearly define the focal question.

"What are the implications for the UK by 2020 if it leaves the EU in 2017 following a "leave" vote in the 2016 referendum?"

Step 3: Write down as many thoughts as possible in response to that question, however odd they may sound. Do not self-censor.

- Overall damage done to UK economy
- Loss of free trade – EU is UK's biggest market
- Businesses (and jobs) might leave the UK
- Investors might leave the UK
- Scotland might secede and UK will fracture (Chinese/Mid-East investment in Scotland = security risk for rest of UK)
- Reduced immigration could be good or bad for the economy
- What happens to cheap EU flights/holidays?
- Political tensions with EU? Implications?
- Freedom from unelected Brussels beaurocrats
- UK regains its sovereignty
- Risk to UK security?
- Avoid EU's goal of increasing integration/control
- Loss of common agricultural policy (good or bad?)
- Less economic regulation (good or bad?)
- Improved trade with Commonwealth etc.
- Schools teach Mandarin instead of French/German
- Lose opportunities to work/study abroad
- No reciprocal free health care when in EU
- E.g. chorizo and BMWs may cost more
- Loss of coordinated anti-terror activities
- Can't make good deals abroad without power of EU bloc
- Big loss of UK world influence – implications?
- Collapse of EU eventually results in WWIII across Europe or invasion from Russia?
- UK saves 0.5% GDP in EU payments
- Closer social/political relations with non-EU countries e.g. China, India, Brazil

Step 4: Arrange responses into appropriate categories in order of importance to the issue of interest. Consider using a mind map to arrange and develop ideas.

See the example mind map below, with responses arranged numerically according to their importance.

Economic

1. Overall damage done to UK economy.
2. Loss of free trade – EU is UK's biggest market.
3. UK saves 0.5% GDP in EU payments.
4. Businesses (and jobs) might leave the UK.
5. Investors might leave the UK.
6. Reduced immigration could be good or bad for the economy.
7. Less economic regulation (good or bad?)
8. Loss of common agricultural policy (good or bad?)
9. Can't make good deals abroad without power of EU bloc.

Political

1. UK regains its sovereignty.
2. Closer social/political relations with non-EU countries e.g. China, India, Brazil.
3. Big loss of UK world influence – implications?
4. Political tensions with EU? Implications?
5. Risk to UK security? E.g. no Europol.
6. Loss of coordinated anti-terror activities.
7. Freedom from unelected Brussels beaurocrats.
8. Avoid EU's goal of increasing integration/control.
9. Scotland might secede and UK will fracture (Chinese/Mid-East investment in Scotland = security risk for rest of UK).
10. Collapse of EU eventually results in WWIII across Europe or invasion from Russia?

UK leaves EU

Social/other

1. Lose opportunities to work/study abroad
2. No reciprocal free health care when in EU
3. Schools teach Mandarin instead of French/German
4. What happens to cheap EU flights/holidays?
5. E.g. chorizo and BMWs may cost more