



National Security Journal

<http://nationalecurityjournal.nz>

Published by:  
Centre for Defence  
and Security Studies,  
Massey University

ISSN: 2703-1926 (print) ISSN: 2703-1934 (online)

---

## Seeing the Futures: Evaluating the Application of Structured Analytic Technique Alternative Futures Analysis

Author: Loren McCarthy

To cite this article: McCarthy, L. Seeing the Futures: Evaluating the Application of Structured Analytic Technique Alternative Futures Analysis. *National Security Journal*. Published 03 November 2024.  
doi: 10.36878/nsj20241103.08

To link to this article: <https://doi.org/10.36878/nsj20241103.08>

View CrossRef data: <http://search.crossref.org/q=10.36878%2Fnsj2024.1103.08>

# SEEING THE FUTURES: EVALUATING THE APPLICATION OF THE STRUCTURED ANALYTIC TECHNIQUE ALTERNATIVE FUTURES ANALYSIS

Loren McCarthy<sup>1</sup>

The Structured Analytic Technique (SAT) Alternative Futures Analysis (AFA) is an imaginative thinking technique used by intelligence analysts in circumstances of uncertainty to generate and explore a range of potential future states. Recent literature on SATs has been critical of their use because of an absence of robust evidence of their value. This research observed focus groups which applied AFA to a specific case study and evaluated their outputs. The results suggest AFA can be used less prescriptively than some of the literature sets out in terms of group size and timeframe, while also supporting the recommended use of a facilitator. An understanding of the purpose of AFA as an imaginative thinking technique, intended to explore future possibilities, is vital in applying the technique, evaluating it, and communicating the results. The potential for mismatch between the criteria suggested to analyse SATs in general and the intent of AFA as a technique, undervalues the technique. A key consideration in evaluating AFA, is the need to ensure the technique is applied as intended. Further, this research identifies that the learning context may influence the application of AFA.

Keywords: Structured Analytic Technique, Alternative Futures Analysis, Intelligence Analysis, Futures Studies, Scenario Generation.

---

<sup>1</sup> Loren McCarthy holds a Master of International Security (intelligence) from the Centre for Defence and Security Studies, Massey University. This article is based on her research dissertation. The author would like to acknowledge the contribution of Dr John Battersby, Senior Fellow at Massey University's Centre for Defence and Security Studies, as supervisor of this research project. Dr Battersby's guidance, expertise, and support has been invaluable. Any correspondence in relation to this article, please contact the Managing Editor, NSJ at [CDSS@Massey.ac.nz](mailto:CDSS@Massey.ac.nz).

## Introduction

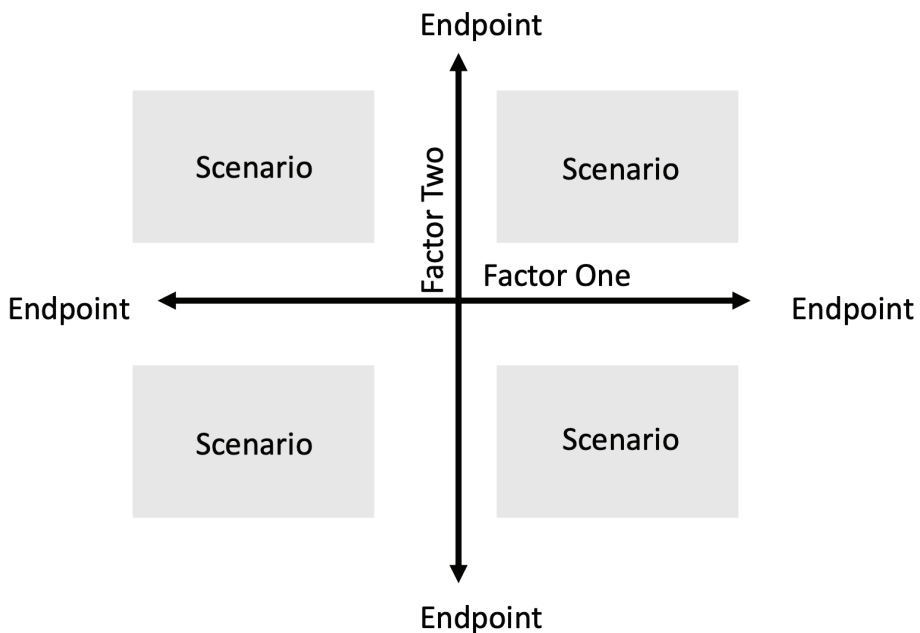
Structured Analytic Techniques (SATs) are the name for a group of methods intended to improve intelligence analysis with the intent of allowing transparency, externalising thought processes, and mitigating bias. Efforts to improve analytic processes are not new; the need for 'alternative analysis', to challenge the user to consider alternatives, was identified in the 1980s. The terror attacks of 11 September 2001 and the intelligence controversy concerning weapons of mass destruction in Iraq of 2002 are considered to have intensified wider efforts to improve analysis. Some of the initial shortfalls identified in these cases were 'failure of imagination', failure to challenge key assumptions, and to consider alternatives. Following these cases, the scope of alternative analysis was expanded to structured analysis.

SATs are intended to provide transparency, instil rigour, and address subjectivity introduced into analytic processes through bias. They are often supported by specific guidance for analysts to follow. The intent is that by following a structured and transparent process, the probability of error introduced by bias is reduced. SATs are also intended to facilitate transparency through illustrating the process behind judgments. They can be used as structured processes to enable collaboration on problems. Grouped by purpose, SATs include diagnostic techniques, which illustrate arguments, assumptions or gaps, contrarian techniques, which challenge thinking, and imaginative thinking techniques, which seek to develop different perspectives or outcomes.

Alternative Futures Analysis (AFA) is an imaginative thinking or scenario technique used in complex situations when the future outcome may be uncertain. AFA requires the user to select the two most uncertain and critical factors in relation to a particular issue, and consider how the interaction of these two factors might shape the future. These factors are somewhat inconsistently referred to as factors, forces, drivers, or uncertainties throughout the literature. Factors are intended to represent 'driving forces' which are considered likely to influence the future. Joseph Nye suggests that estimators should not attempt to predict the future, but should aim to present alternative futures, and the signposts which might identify alternative paths. In this sense, AFA is not intended to predict what the future will be, but instead to imagine a range of scenarios that explore what the future *could be*. Following this, each scenario can be mapped back to the present, and indicators can be developed which might be used to track whether a particular future may be developing. The technique is believed to have been developed by military planners, but is also used in business. The work of Herman Kahn and the RAND Corporation in the 1950's in scenario planning is considered to have paved the way for the consideration of hypothetical futures. The most well-known use is the frequently cited 1970s use of the technique by the Royal Dutch/Shell corporation to prepare for future oil crises.

The method of AFA is described across literature, as applied across a wide range of disciplines. In the policy realm the UK Government Office for Science’s 2017 “The Futures Toolkit” refers to it as ‘Axes of Uncertainty’. Randolph Pherson also presented a variation of AFA called Multiple Scenarios Generation, which essentially uses a similar method to AFA, but facilitates the use of more than two factors. The method of AFA, as presented in United States Government’s “A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis”, is summarised as follows:

- Develop the issue being considered
- Brainstorm relevant factors that affect the issue
- Select the two most critical and uncertain factors to turn into axes
- Establish relevant endpoints
- Create a matrix from the two selected axes
- In the resulting matrix (*Figure 1*), develop stories describing the futures created by the combination of the two factors. Develop indicators for these futures.



*Figure 1: Example AFA matrix with identified factors, end points, and scenarios.*

The value of AFA is in its ability to manage and explore uncertainty. It is considered a divergent thinking technique that describes multiple outcomes for consideration, rather than predicting a single outcome.<sup>1</sup> AFA assists consideration of otherwise surprising developments, and allows for the development of indicators to monitor signs that an identified future possibility might be likely to occur.<sup>2</sup> The value is ultimately the ability to consider how decisions or policy may perform in each of the futures, and develop agile or flexible approaches.<sup>3</sup>

The term 'alternative futures' is also used in the literature in a general sense to describe the idea that 'the future is plural'.<sup>4</sup> There are multiple methods of developing alternative futures.<sup>5</sup> Underpinning AFA is the concept that in theory multiple possible futures exist.<sup>6</sup> Thus the idea of alternative futures shapes the understanding of the specific method of AFA through facilitating the exploration of multiple futures beyond that considered most likely.

The limitations of AFA are not particularly clear in the methods. The resources required to facilitate AFA can be significant, with multiple sources suggesting it is labour intensive, requiring considerable time in hours or days, and the involvement of experts in addition to numbers of analysts.<sup>7</sup> The primer also refers to being open to engaging in the process, which is described as more 'free-wheeling'.<sup>8</sup> Pherson is perhaps most prescriptive in presenting the methodology, stating AFA should only be used for 'simple situations', over a longer time frame (three to 10 years).<sup>9</sup> The participation of 24-40 participants is recommended, with facilitation by a team (of four).<sup>10</sup> Pherson categorically states the technique should not be used if there are more than two critical factors.<sup>11</sup> Ian Wing specifically outlines problems to look out for when using AFA.<sup>12</sup> These include scenarios which could be ambiguous, and that unexpected developments could influence all of the futures. Wing also suggests that the concept of extrapolating trends itself relies overly on historic examples, meaning 'wild cards' (defined in the context of futures studies by Elina Hiltunen as 'rapid (and in that sense surprising) events that have vast consequences...')<sup>13</sup> may be discounted. Wing however appears confident that all of these potential weaknesses can be addressed.<sup>14</sup>

## Evaluation of SATs

Literature on the use of SATs critically highlights the absence of robust analysis of their efficacy.<sup>15</sup> Many SATs are considered not to have undergone sufficient academic scrutiny or been rigorously tested in a way that would demonstrate their value.<sup>16</sup> This limits the ability to comment on whether SATs in fact help or hinder the formulation and delivery of accurate intelligence assessments.<sup>17</sup> Recent criticism observes that the intelligence literature appears to simply assume that they work as intended, despite rarely being subject to scrutiny or critical analysis.<sup>18</sup> Chang, Berdini, Mandel, and Tetlock contend that SATs hold much promise, but without robust and demonstrated efficacy,

risk lack of integration and adoption.<sup>19</sup> Stephen Coulthart comments on the wealth of case knowledge relating to intelligence practice, juxtaposed with the lack of scientifically validated research.<sup>20</sup> A 2016 RAND Corporation study suggested SAT use was increasing, but suggested at that time that the ‘U.S. Intelligence Community does not systematically evaluate the effectiveness of structured analytic techniques.’<sup>21</sup>

Suggested methods of evaluating SATs refer to the scientific gold standard of randomised control trials, and appear to focus predominantly on accuracy of assessments.<sup>22</sup> Chang *et al.*, also mention logical validity, bias bipolarity, noise, reliability, validity, accuracy, internal coherence, and quality of explanations could be considerations in the examination of the efficacy of any SAT.<sup>23</sup> Chang *et al.*, do acknowledge that to evaluate accuracy it might be suitable to check whether the ‘correct’ hypothesis was ‘on the radar’ throughout the process, and to determine the reasons it was undervalued.<sup>24</sup> A 2022 examination of the SAT Cone of Plausibility (COP) which evaluated scenario quality concluded that scenarios produced using this technique were rated high or medium quality based on the five criteria used; completeness, plausibility, context, coherence, and order effects.<sup>25</sup> The study suggested additional research should be completed to address whether the technique reduced hindsight bias, was useful as a forecasting technique, and supported decision making.<sup>26</sup>

The RAND study considered that little effort had been made to assess the efficacy of SATs generally, as well as specific SATs.<sup>27</sup> The study conducted a preliminary review of a sample of publications, finding that those which used SATs tended to consider a broader range of outcomes or implications than others.<sup>28</sup> The study proposed a multi part approach to evaluating SATs, made up of interviews and reviews of work.<sup>29</sup> It also mentions the use of controlled experiments to test the application of SATs.<sup>30</sup> The study listed relevant research at the time, highlighting mixed results, and in some cases limited relevance to conditions in the intelligence community.<sup>31</sup>

A more recent 2019 study which examined the diagnostic technique Analysis of Competing Hypotheses (ACH) found mixed evidence with regards to the ability of the technique to mitigate confirmation bias, and suggested its use may in fact increase inconsistency of judgments.<sup>32</sup> It also noted that most individuals in the group instructed to use ACH departed from at least one of the steps in the method. The study also observed at least one instruction manual explained ACH incorrectly, with the result that analysts and their organisations may believe they are applying ACH when they are not necessarily using it as intended.<sup>33</sup> While an important finding, this potentially speaks to a limitation of this and other studies seeking to evaluate SATs, in that the technique or method itself cannot necessarily be empirically tested if it is not being used correctly or as intended. The 2019 study, which used a trained ACH group, and an untrained control group, observed the majority of the non-ACH group re-formatting their data, with over half the group’s participants arranging it into an ‘ACH-style matrix with hypotheses as

columns and evidence as rows.<sup>34</sup> This suggests the majority of the non-ACH group applied some form of structured technique, in which case, ACH was not necessarily being compared with a suitable control, but with another technique (perhaps even itself).

While there are some studies which have sought to evaluate AFA, there has also been recent criticism relating to these. In a study of the twelve techniques in “A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis”,<sup>35</sup> Coulthart comments that AFA (along with Devils Advocacy and Brainstorming) have longer research histories and a number of evaluative studies, many of which come from outside of security or intelligence disciplines.<sup>36</sup> Almost all of the 17 evaluative studies of AFA Coulthart analysed were considered to be ‘highly effective’, but the studies had ‘reduced credibility’ of evidence.<sup>37</sup> Coulthart suggests this warrants low confidence in the findings of these studies.<sup>38</sup> The criteria for being considered effective in his study was improving ‘rigour, accuracy, or another outcome’ without negatively affecting any other outcome.<sup>39</sup> Further specific challenges with SATs have also been subject of comment. Chang *et al.*, suggest the targeted bias in AFA is status quo bias, which refers to the ‘no change’ hypothesis being given more weight than warranted.<sup>40</sup> Theoretically, techniques aimed at mitigating this bias should challenge the analyst to consider unanticipated change.<sup>41</sup> Chang *et al.*, suggest targeting this bias then creates the risk of activating the opposing bias of base-rate neglect (or incoherent probabilities<sup>42</sup>), which could result in probabilities being inflated or deflated, and therefore being judged as more or less probable than evidenced.<sup>43</sup> This existence of opposing bias, which could potentially be activated, is referred to as bias bipolarity.<sup>44</sup> A second challenge posed by Chang *et al.*, is the understanding of whether problem decomposition reduces or in fact adds to ‘noise’.<sup>45</sup> This refers to the idea that in decomposing a problem into component parts for analysis, subjectivity may be introduced and carried forward into subsequent analysis, introducing additional ‘noise’ to the process.<sup>46</sup>

Difficulties with evaluating SATs are explored across the current literature. The RAND study also suggested the difficulties with evaluation are part of the reason why SATs have not been evaluated in detail.<sup>47</sup> Stephen Marrin concluded in 2012 that the criteria often applied to assessing the quality of intelligence analysis each had limitations, and there was no consensus on which criteria should be used, and further, no consensus on the purpose of intelligence analysis.<sup>48</sup> Attempts at evaluation have also been criticised for not accurately reflecting conditions in which the techniques might be used.<sup>49</sup> As the concept of developing scenarios could be achieved through the use of a number of techniques,<sup>50</sup> identifying that AFA was actually used is vital to assessing it specifically. The RAND study suggested one of the difficulties in evaluating SATs is that it can be unclear if a SAT has been used unless this is explicitly stated.<sup>51</sup>

Various criteria have been applied or suggested for use in the evaluation of SATs, or of completed analysis. In their study the RAND Corporation used the criteria set out in the Intelligence Community Directive (ICD) 203 (2007) as issued by the US Office of

the Director of National Intelligence (ODNI).<sup>52</sup> These tradecraft standards are however intended for ‘finished intelligence’.<sup>53</sup> This highlights another challenge in evaluating SATs, namely that of evaluating the effect of the SAT use in isolation of other factors which might impact analysis. Marrin also discusses criteria that apply to analysis overall, but not specifically to SATs.<sup>54</sup> Both Marrin and the RAND study also discuss the concept of decision maker evaluation of analysis and the challenges associated with this.<sup>55</sup> It remains unclear if there is a specific consensus on how to evaluate SATs, and if they should be evaluated on their own, or as part of a finished product. Marrin in particular highlights specific issues with the commonly suggested ‘accuracy’ criteria, namely the use of ‘qualifiers’ to communicate uncertainty, and the fact that analysis could influence outcomes, essentially making itself inaccurate.<sup>56</sup> Richards Heuer also addresses the ‘accuracy’ criteria, arguing that judgements are often conditional on other factors, use probability, the difficulty in conducting experiments that are reflective of technique use in the intelligence community, and the influence of personal experience with SATs on opinions of them.<sup>57</sup> Heuer argues instead SATs should be evaluated based on whether they achieve their purpose, and if there is not a better alternative.<sup>58</sup>

Some of the literature relating to the assessment of the efficacy of SATs is contradictory to the current understanding of the intent of AFA. As discussed scenarios produced by AFA are not intended to provide accurate predictions, but rather they aim to imagine a range of possible future states.<sup>59</sup> Yet the standard criteria suggested, including by Heuer and Pherson,<sup>60</sup> to examine the efficacy of SATs is generally based on the accuracy of the technique.<sup>61</sup> Wing reiterates that AFA is not intended to predict the future, but should be used to encourage original thinking.<sup>62</sup> Wing further states that ‘[t]his is not a problem as long as all those involved in the use of AF[A] understand their intended use.’<sup>63</sup> Assessing AFA based on accuracy criteria is potentially counterproductive to the intent of the technique.

The current state of the literature relating to the empirical evaluation of SATs, potential criteria for testing, and in particular those relevant to AFA, does little to robustly validate or assess the technique. The potential for mismatch between the criteria suggested for use to analyse SATs in general and the intent of AFA as a technique risks that the technique might be under or overvalued. The limitations and strengths of AFA appear to lack robust analysis and without a solid grounding in the theory of futures analysis the value of AFA risks being miscommunicated.

## Research Methods

This study consisted of five focus groups (five to six participants in each group) which applied AFA to a case study. Initial findings were reviewed following the first set of four focus groups, and a modified approach was applied to a second single focus group. In each case the researcher presented a short tutorial on the intent and method of the SAT AFA. The groups were instructed to use the AFA technique as described in United States



Government's Tradecraft Primer.<sup>64</sup> The application of the technique was observed, and the outputs evaluated based on criteria from the literature review. Thematic analysis was conducted on survey responses (16 from round one, five from round two) following application of the technique to identify key themes in the understanding and application of the technique.

The first set of focus groups were instructed to use the method set out in the Tradecraft Primer,<sup>65</sup> and provided with a copy of the method. Two of these groups considered the SAT scenario for a longer-term time frame, and the other two considered a more immediate future. The case study used was the acts of terrorism by Ananda Marga (AM) in Australia and New Zealand in the 1970's, prior to the bombing of the Hilton Hotel in Sydney in 1978, as researched by John Battersby.<sup>66</sup> The participants were not familiar with Battersby's research prior to the presentation of the case study and did not know of the February 1978 Hilton bombing. Being provided with only background detail known prior to the bombing, the focus groups were asked to use AFA to provide an estimate of the likely risks associated with AM activity. These groups were observed, but were not actively facilitated.

Initial findings from the first focus groups were integrated into an adapted approach for a second separate and subsequent focus group to further explore some of the themes derived from the initial survey responses, in particular some unanticipated findings, together with an attempt to address identified limitations from the first groups. The second case study used asked the focus group to consider the potential future risks and impacts of the use of Unmanned Aerial Vehicles (UAV's or drones) on the New Zealand policing environment. In addition to the method presented by the researcher, the group also received a presentation from a subject matter expert on drones, and was introduced to the idea that the process could be repeated using more than two factors. The second focus group was actively facilitated, as is suggested in the literature.<sup>67</sup>

### *Criteria*

The focus group outputs were evaluated against a set of five key criteria derived from the literature review, as listed below. Some of these criteria were not able to be assessed, for example, not knowing the outcome of the case study used for the second focus group prevented the measurement of its accuracy. There were also several unanticipated findings which impacted the use of the criteria (discussed below). The criteria used were:

**Accuracy:** if the outcome (where known) of the case study was included in the AFA matrix, it was considered to have encompassed the 'correct' answer. Consideration was also given to whether the outputs captured a range of outcomes (completeness).

**Validity:** if the scenarios generated were logically based on the combination of factors considered, it was considered to be valid for the purposes of this research.

**Decomposition and noise** (as in Chang *et al.*,<sup>68</sup>): this measure was used to identify if the way the problem was broken down, specifically through the selection of factors and end points, was consistent across the groups. This measure was not able to be applied for the second focus group, as there was only one group involved.

**Reliability:** this measure was used across different groups to identify if the scenarios generated by different focus groups were consistent with one another. This measure was also not able to be applied for the second focus group, as there was only one group involved.

**Evidence of bias mitigation:** if the scenarios generated showed sufficient range to indicate multiple scenarios were considered, status quo bias was considered to have been addressed.

### Reflexive Thematic Analysis

Thematic analysis was undertaken on the survey responses, to identify themes in the understanding and application of AFA, and for comparison with observations. The methodology used to facilitate this was the six step reflexive thematic analysis methodology set out by Braun and Clarke.<sup>69</sup> Reflexive thematic analysis was selected as a methodology as it allows scope to incorporate and recognise the subjectivity of the researcher.<sup>70</sup> This aligned with the somewhat ethnographic component of the research design in which the researcher was involved in teaching the technique to the focus groups, and then observing their application of it. The overall participation in the research was five focus groups of approximately five to six participants each, and 21 survey responses.

### Limitations

The use of focus groups rather than individual application of the technique was intended to address the criticism that some experiments have not sufficiently replicated the conditions in which SATs are used in the intelligence community, namely that they have assessed technique use by individuals rather than groups.<sup>71</sup> It should be acknowledged that the trade off in addressing this criticism is introducing an element of group dynamics which may not be able to be controlled for. However, as AFA is designed as a group technique this risk would always be present.

An acknowledged limitation of the study design is the lack of a specific control group. As noted in Chang *et al.*, and the RAND study, controlled experiments could be used to test SATs,<sup>72</sup> and could be considered the scientific gold standard for evaluation.<sup>73</sup> The difficulty of conducting these experiments, particularly in light of complicating factors such as group dynamics, relevance to real life use, and the need to ensure the method is being followed correctly is apparent in the extant literature.

## Results

The evaluation of decomposition and noise, which refers to consistency in the breaking down of a problem, as suggested in Chang *et al.*,<sup>74</sup> was intended to be evaluated through comparing the factors and end points from the first four focus groups against one another in order to identify similarities or differences in the way the problem was broken down. This was impacted by an unanticipated finding. The methodology used in the Tradecraft Primer presented an example framework, which considered two specific factors (*Table 1*).<sup>75</sup> Some of the focus groups started their AFA framework with these same or similar factors, in some cases word-for-word on their axes. Only one of the longer-term focus groups selected notably different axes.

Group	Axis One	Axis Two
Eg.	Intended Impact (selective/broad)	Sophistication of weapon (low/high)
1	Scope of impact (low impact/high impact)	Degree of violence (low violence/high violence)
2	Intended Impact (selective/broad)	Capability (low/high)
3	Intended Impact (selective/broad)	Method (decreasing violence/increasing violence)
4	Escalation (legitimisation/radicalisation)	Expansion of group (low/high)

*Table 1: Focus Group Axes Selection Compared to Example.*

Due to this potential influence from the example in the way the groups selected their axes, it would be inappropriate to comment on how similar their selection of axes was. This was factored into the way the final focus group (considering drones) was run. During the teaching component of the final focus group, the researcher presented a different, trivial and unrelated example, using factors which it would be very difficult to relate to the case study being used.

This also impacted the evaluation of the reliability or consistency. The groups with similar axes did have similar scenarios and consistent placement of these within the matrix. For example, three of four groups specifically included a bombing attack in their top

right-hand quadrant, and three of the four had protest action in their bottom left quadrant. In each case, the group that did not have bombing or protest in the same quadrant as the other groups did, still had each on their matrix, just differently placed. The influence from the example on the selection of axes likely carries through into the generation of scenarios, and thus it would not be appropriate to comment on how consistent the use of the technique was.

With regards to accuracy or completeness, for the first set of focus groups all four groups had a scenario or part of a scenario that could have been considered consistent with the known outcome of the case study, in some cases, including specific relevant detail. The case study developed into the bombing of the Hilton Hotel in Sydney in 1978 where the Commonwealth Heads of Government Regional Meeting was being held.<sup>76</sup> All four groups had a scenario involving bombing. However, the two short-term focus groups in particular included specific detail, with one listing 'Bombing at diplomatic event or Indian High Commission' and the other 'Bombing- Indian targets, domestic'. Considering completeness rather than accuracy, most groups included a 'nothing happens' or low risk scenario, through to an 'extreme' or high-risk scenario, and several scaled scenarios in between.

Scenario quality may have been limited by time. Some groups developed information beyond a headline for the scenario, and all developed a list of possible indicators that might be precursors to the scenarios eventuating. Scenarios were well reasoned, and they made sense based on the factors selected. As the groups used bullet points or lists to describe the characteristics of each scenario, scenario quality was not able to be assessed. The groups did not produce finished narratives as AFA is intended to achieve,<sup>77</sup> but rather identified characteristics of each scenario. Some groups clearly created four scenarios, one per quadrant, while other groups appeared to have multiple alternative scenarios within the same quadrant. For example, one group had 'hijack airplane', 'bombing' and 'kidnapping' in the same quadrant. The guidance on this is somewhat unclear in the Primer, with the example using bullet points and having multiple ideas in each quadrant.<sup>78</sup> The scenarios could not be judged on coherency where coherence was not necessarily instructed in the Primer.

With regards to evidence of bias mitigation, the intention of the exercise was to explore how AFA could mitigate confirmation bias by attempting to capture the ideas of participants before and after application of the technique. However, the 'before state' of the first groups was unable to be sufficiently captured, so this was not possible to measure. In the second focus group, the question was altered to reflect this better and generated some appropriate responses. The initially generated ideas appeared to fit predominantly within one or two quadrants in the final matrix, indicating that a wider range of options were generated following the use of AFA than before it. Due to the small number of appropriate responses, this would require further testing to be confidently validated.

A second unanticipated finding was noted in the focus group outputs and observation in that at least one group appeared to have initially brainstormed potential outcomes of the scenario (as opposed to the factors influencing those outcomes), essentially bypassing the analytical process. This group had carried out a brainstorm of ‘likely activities’, which were then reflected in their matrix. In doing so, they may have essentially reverse engineered the AFA so that they decided on the outcomes first and then made them fit into the framework.

### *Thematic Analysis*

The themes and sub-themes identified in the post focus group survey responses are described below, with selected responses included for illustration.

#### *Understanding the purpose of AFA*

This theme presented as sometimes conflicting sub-themes relating to AFA as an idea generation technique for the purpose of generating a range of potential futures, and expectations of providing specific outcomes, or not being valuable if it did not offer predictions. These conflicting sub-themes sometimes came from the same participant, or even within the same answer.

“While the tool potentially allows the exploration of multiple alternatives...you end up with 4+ scenarios but zero indication of what might be more likely.” [AH, Round 1, Focus Group 2]

Responses such as these appeared to recognise the idea generation aspect of the technique, while also conveying an expectation that a specific answer was desirable and achievable. The second-round focus group also contained themes relating to the understanding of AFA, and while predominantly relating to the sub-theme of AFA as an idea generation technique for the purpose of generating a range of potential futures, and a new sub-theme relating to the value of indicators and preparedness, at least one response related to the concept of a ‘most likely’ scenario.

“I would advise what the most likely scenario is and how to identify indicators of change to prevent it or intervene early.” [BC, Round 2, Focus Group 6]

“...there are many possible futures that will be influenced by a number of factors. The indicators identified will need to be monitored to identify what will happen and preparedness will need to take place.” [BE, Round 2, Focus Group 6]

Other responses centred on beliefs about decision maker expectations around what would be helpful or considered useful from their perspective.

“...it did leave us with the ‘so what?’ question about which outcome is the most likely. That is probably what policy makers are going to want to know.” [AN, Round 1, Focus Group 4]

“I can see the benefit for decision makers. Rather than being presented with the most likely scenario, [AFA] offers a range of scenarios for them to consider. Gives them a heads up and no surprises. Opportunity to take action to lessen risk or place flags on indicators...” [AR, Round 1, Focus Group 3]

There were also conflicting ideas within beliefs about decision maker expectations, with some responses indicating the belief the technique would not be helpful for decision makers, and others identifying how it would be beneficial. The second round focus group appeared clearer on the value added for decision makers.

“I would outline the four scenarios....to add value for the client so they [are] informed as to our thinking and are open-minded that one clear scenario is not certain.” [BD, Round 2, Focus Group 6]

A sub-theme also captured under understanding the purpose of AFA was the need for experience with the technique, or facilitation. This was considered a sub-theme to understanding purpose as it frequently related to lack of certainty around how to use the technique.

“...don’t know if I would necessarily do one again unless facilitated by someone confident in the application of this SAT.” [AE, Round 1, Focus Group 1]

This sub-theme was also present in the responses from the second focus group round who did receive facilitation, but was phrased positively in terms of the value of having a facilitator present.

“...the facilitator help[ed] in terms of gently steering us back on course or consider other perspectives. It really made me consider the value of having a facilitator.” [BC, Round 2, Focus Group 6]

### *Axes selection shapes outcomes*

Axes selection was considered difficult and important. The second round focus group also had this theme.

“Critical to get this [deciding on the force axes] right I think as if you don’t your potential outcomes won’t be valid.” [AN, Round 1, Focus Group 4]

“The importance of these [axes] isn’t really explicitly stated – and changing them could potentially provide completely different results.” [AH, Round 1, Focus Group 2]

Some responses considered that axes selection was limiting, or made the outcomes not exhaustive. Some mentioned that ‘wild cards’ were not able to be generated using the technique. Others suggested that this could be solved by simply selecting more than two axes and repeating the technique multiple times.

“...it was difficult to frame the axes, and this could mean that some events are left off if the axes were not the most efficient.” [AD, Round 1, Focus Group 1]

“... the axis used can put limitations on the outcomes. Maybe this tool needs to be used several times for the same case study if there are a range of options for the axis so all possibilities can be explored?” [AG, Round 1, Focus Group 2]

The second round focus group was also asked to comment on the potential for inclusion of ‘wild card’ scenarios. Some were suggested, while other responses suggested multiple rounds of AFA would capture more ‘wild card’ scenarios.

“...I think doing a few crosses of critical factors can help this [suggesting ‘wild card’ scenarios]” [BB, Round 2, Focus Group 6]

“...I feel the scope of the four scenarios ... covers the spectrum for a decision maker.” [BD, Round 2, Focus Group 6]

A less common theme, but nonetheless significant in terms of its impact on the use of the technique, was the use of outcomes to shape or identify the factors used in the AFA matrix. This linked to the idea that axes selection was difficult. This was not present in the second round focus group, who were facilitated.

“Seemed easier to think of examples before thinking of axes labels.” [AQ, Round 1, Focus Group 5]<sup>9</sup>

### *Necessary for good AFA*

Multiple sub-themes related to the overall theme around what was necessary to undertake a good AFA. These included the participation of a subject matter expert (SME) in the context, as distinct from expertise in AFA itself, time, and the value of completing the technique as a group.

“The results are only as good as your subject matter knowledge...” [AE, Round 1, Focus Group 1]

“Time pressure impacted on discussion and selection of the axis.” [AM, Round 1, Focus Group 4]

“...even in a short time with limited info you can have some sound discussions and generate sound ideas.” [AN, Round 1, Focus Group 4]

The second round focus group also had themes relating to the involvement of a SME and the application of the technique by a group. The involvement of the SME was seen to be valuable. The time sub-theme was not evident in responses as the group was not limited for time.

“... [reflecting on what was helpful] ...the information about the subject matter (i.e. drones) it gave context to what we would later discuss.” [BA, Round 2, Focus Group 6]

#### *Link with other techniques*

Another theme identified was reference to or incorporation of other techniques to support the AFA. Frequently mentioned was brainstorming, including the concept that AFA was essentially a form of structured brainstorming. PESTLE<sup>2</sup> was also used by some as part of identifying factors. Indicators were used by all groups in line with the instructions in the method.

“As a standalone tool it doesn’t seem particularly thorough.” [AC, Round 1, Focus Group 1]

“..inclusion of structured ways of thinking in the brainstorm could be helpful (PESTELOM).” [BB, Round 2, Focus Group 6]

## **Concluding Discussion**

### *Considerations for use of AFA*

The results of the focus groups suggest AFA can be used less prescriptively than some of the literature suggests. Groups of five to six, using shorter time frames than suggested, were able to produce meaningful scenarios and indicators for these. As Dhami, Wicke, and Onkal addressed in their study of COP, while the study of the future often suggests longer time horizons be used,<sup>80</sup> it is partly dependant on the issue being considered,<sup>81</sup> and in the security domain futures research can focus on both longer and shorter term time frames.<sup>82</sup> The Cone of Plausibility study also used single participants, while Pher-son prescribes three to twelve participants with a facilitator for this technique.<sup>83</sup> AFA was also able to be applied over a matter of hours. This suggests that AFA, and other

<sup>2</sup> Political, Economic, Social, Technological, Legal and Environmental.



futures techniques, may be more flexible than generally considered. Future research might consider comparing group sizes, or the use of the technique by single individuals, and over varying time horizons.

Nonetheless, the focus group results also strongly supported some of the recommendations in the literature for the application of AFA. In particular, the use of an effective facilitator was a significant theme amongst groups, both in absence and when present. While AFA may be more flexible than anticipated, and methodologically straight forward, it is not a simple technique. A facilitator with a keen understanding of the theory behind futures research, and experience in the use of AFA specifically may be necessary to ensure the technique is both used and communicated appropriately.

There is perhaps an unnecessary differentiation between AFA and Multiple Scenarios Generation in some literature. While Pherson specifically states AFA should be used in simple situations when there are only two factors, and Multiple Scenario Generation should be used in more complex situations when there are more than two factors,<sup>84</sup> other authors do not make this distinction. Wing suggests creating a three-dimensional scenario space in which additional axes are introduced.<sup>85</sup> A common criticism of AFA was that the factor selection limited the outcomes, leading to the alternative futures hypothesised not being exhaustive. This might be solved by removing the limitation of only allowing two factors. The viability of this is an area requiring further research.

Understanding the purpose of AFA is a key aspect of using, communicating, and evaluating it as a technique. The expectation that AFA should provide predictions, or was not valuable to a decision maker if it did not, contradicts the intent of the technique. A clear understanding of its purpose as an imaginative thinking technique looking to explore the range of potential future states in circumstances of uncertainty is therefore vital in applying the technique and communicating the results.

This purpose links in with the idea of AFA as a technique intended to mitigate status quo bias. Theoretically, the technique should do so in its design, by imagining more than one future. The analyst using AFA has to think beyond a simple extrapolation of the status quo and is forced to imagine different futures. The idea that this might activate the opposing bias of base-rate neglect or incoherent probabilities, as explored by Chang *et al.*,<sup>86</sup> is not necessarily relevant in a technique that does not pretend to be able to assign probabilities to the alternate futures it develops. AFA simply imagines the futures, it does not claim to predict which will occur. It is perhaps not in the use of the technique, but in the misuse, that base-rate neglect or incoherent probabilities may become troublesome. These opposing biases, status quo and base-rate neglect or incoherent probabilities, are perhaps more applicable to finished analysis to which AFA has contributed, and then indicators have been applied and monitored, in which case it might be valuable to examine how these might present in analysis which has used AFA, compared to analysis which has not.

While AFA in itself does not assign probabilities to futures, it does link into a secondary process which suggests it is able to do so - the generation of indicators.<sup>87</sup> This should perhaps be considered an entirely different and distinct technique from AFA itself, and indeed is covered separately in the Tradecraft Primer as the 'Indicators or Signposts of Change' SAT.<sup>88</sup> This technique suggests it can be used to '[i]dentify the most likely or most correct hypotheses or scenarios, based on the number of changed indicators that are observed'.<sup>89</sup> AFA alone suggests only a range of potential futures that could occur, it is the application of indicators to these scenarios that might be used to identify if a potential future is beginning to develop. This, and the linkages with other techniques made by the focus groups, suggests the effective use of one technique may in practice necessitate the use of several interlinked techniques. If this is the case, it is important users understand which technique is intended to do what.

It could also be argued that AFA has value even without moving on to identifying which future is emerging. The Tradecraft Primer discusses the value of AFA for policy makers to design flexible policy that could work in multiple futures, thereby focussing on preparedness for a range of futures in the face of uncertainty.<sup>90</sup> In this sense, AFA could be seen as allowing the decision maker to prepare an approach to survive across a range of futures, rather than betting on one. As one first round participant noted - the technique "[g]ives them a heads up and no surprises" essentially performing the task AFA was designed for.

The themes derived from the focus group responses also raise the question of whether consistency is a desirable characteristic of a divergent thinking technique such as AFA,<sup>91</sup> and thus whether it is an appropriate criterion for evaluation. A repeated theme in the first round of focus groups who were limited to two factors, was that factor selection was limiting, and influenced the scope of outcomes. This suggests they saw value in using more and different factors to produce more and varied futures than could be accommodated within a single matrix. This challenges the idea that consistency is desirable in imaginative thinking SATs, where the purpose is to imagine different outcomes. The use of multiple AFAs with interchanged factors could overcome this problem, as well as displaying the redundancy of consistency as an assessment criterion.

The unanticipated application of the technique essentially backwards by at least one focus group also speaks to the importance of viewing and using SATs as a process rather than an outcome. Where the completed matrix is seen as the end point of the technique, there is a risk that those using it will simply reverse engineer it, coming up with scenarios first, and fitting them into the matrix, rather than using the matrix to consider how different combinations of forces could shape different scenarios. The SAT itself is not the outcome, but a technique to be used as part of the analytical process. Commenting on the subjectivity involved in the interpretation of inputs, Chang *et al.*, describes '[t]he SAT process becomes an end in itself, dressing up subjective judgements in a cloak

of objectivity.<sup>92</sup> More broadly, it may raise the question of whether SATs are understood and used as part of a process to support analysis, or considered an output in themselves by those using them, and if so, whether they are engaged with meaningfully.

The application of AFA raised interesting contradictions with respect to the generation of ‘wild cards’. ‘Wild card’ is a term used throughout literature to refer to surprising events, low probability scenarios, and weak signals.<sup>93</sup> The primer suggests part of the value-add of AFA is in helping to consider ‘wild cards’.<sup>94</sup> Wing meanwhile highlights the risk of discounting ‘wild cards’ as a potential pitfall associated with AFA, and associates ‘wild cards’ instead with ‘future backward scenario development’, as opposed to future forward techniques such as AFA.<sup>95</sup> There were also mixed responses in the focus groups around whether the technique was able to capture ‘wild cards’. This does not appear to be resolved in the literature, and in fact, Hiltunen suggests ‘wild cards’ are not well defined in futures research,<sup>96</sup> which may be at the centre of this contradiction. Whether AFA can adequately capture ‘wild cards’ is perhaps dependent on how analysts define the term.

## Conclusion

### *Lessons for Teaching and Using AFA*

A significant and unanticipated result from the focus groups was the influence the example had on the subsequent use of the technique. For some of the participants, this was their first exposure to AFA, so the example in the primer may have been their only example of the technique being used. The influence from the example manifested not only in the selection of axes, but also in the way the scenarios were developed, using bullet points. This suggests the context in which users learn SATs is important, and influential on their subsequent use of the technique. The learning context may shape the application in future, and careful consideration should be given to the selection of examples in learning and teaching material. Those who have used the technique before might benefit from being mindful of previous use and considering the need to critically consider the factors each time rather than defaulting to previously used or demonstrated axes. Further, this unanticipated result limited the ability to assess a number of the evaluative criteria and as such needs to be kept in mind when designing evaluative studies of SATs.

### *Evaluating AFA and SATs*

A key consideration in evaluating a SAT, as evidenced in this research, is the need to ensure the technique is applied correctly. If it is not, rather than evaluating the SAT itself, there is a risk that the research will in fact be evaluating some other aspect of the process, multiple techniques, the users understanding of the technique, or the wrong technique entirely. For example, at least one of the focus groups brainstormed outcomes rather than factors, and then put these back into an AFA framework. It could be argued

that they did not actually apply AFA, but rather a brainstorming technique, which was then presented visually in an AFA format. Therefore, if the method is not followed correctly, the research risks attributing the success or failure of the analysis to a particular technique, when in fact that technique was not used. The previously mentioned study on ACH noted a tendency even in trained analysts to deviate from the method, as well as identifying material which incorrectly described the methodology.<sup>97</sup> The belief that the technique has been applied, when it has not, is dangerous both in terms of evaluating technique use, but also in applying the technique. Chang *et al.*, argue '[i]f an analyst goes through the technique's motions without significant challenge to their externalized thinking, SATs may provide only a veneer of analytic legitimacy without true improvements to analytic quality.'<sup>98</sup> This is the risk of misunderstanding the technique and how to apply it.

Heuer's suggestion that changing behaviour relating to SATs may be made difficult by resistance and criticism from analysts<sup>99</sup> was reflected in part by some of the survey responses. Participants were indeed critical of what they perceived as shortfalls of the technique, commenting on limitations and expectations. Rather than an obstacle to overcome, however, this is actually an encouraging aspect for the application of SATs. It is not a mark against AFA that participants approached the technique critically, rather than using it blindly. This speaks to the importance of understanding the purpose and limitations of the technique, and recognising what it can and cannot be expected to do.

The purpose of AFA needs to be taken into account when considering evaluating it. It may not be appropriate to use a standardised set of criteria to evaluate all SATs. For example, accuracy is not necessarily an appropriate measure for AFA, as in itself it is not intended to produce predictions. Perhaps accuracy of conclusions stemming from AFA, rather than AFA itself might be appropriate. Potentially more suited to AFA itself would be completeness. This highlights a risk that AFA, and other SATs, could be inaccurately valued if assessed based on criteria not aligned to their purpose.

- 1 United States Government, "A Tradecraft Primer", 34.
- 2 United States Government, "A Tradecraft Primer", 36.
- 3 United States Government, "A Tradecraft Primer", 36.
- 4 Peter Schwartz, author of *The Art of the Long View: Planning for the Future in an Uncertain World* (New York: Doubleday, 1996), quoted in United States Government, "A Tradecraft Primer", 34.
- 5 United States Government, "A Tradecraft Primer", 36.
- 6 Wing, "Short Introduction," 18.
- 7 United States Government, "A Tradecraft Primer", 34; Pherson, "Nine Techniques", 6.
- 8 United States Government, "A Tradecraft Primer", 34.
- 9 Pherson, "Nine Techniques", 2.
- 10 Pherson, "Nine Techniques", 5.
- 11 Pherson, "Nine Techniques", 7.
- 12 Wing, "Short Introduction," 21.
- 13 Elina Hiltunen, "Was it a wild card or just our blindness to gradual change," *Journal of Futures studies* 11, no. 2 (2006): 66.
- 14 Wing, "Short Introduction," 21.
- 15 Chang et al., "Restructuring Structured Analytic", 337.
- 16 Chang et al., "Restructuring Structured Analytic", 337, 349-350.
- 17 Chang et al., "Restructuring Structured Analytic", 337.
- 18 Chang et al., "Restructuring Structured Analytic", 345.
- 19 Chang et al., "Restructuring Structured Analytic", 350.
- 20 Coulthart, "Evidence-Based Evaluation", 368.
- 21 Artner, Girven, and Bruce, "Assessing the Value", 1.
- 22 Chang et al., "Restructuring Structured Analytic", 347, 345-346.
- 23 Chang et al., "Restructuring Structured Analytic", 337-356.
- 24 Chang et al., "Restructuring Structured Analytic", 348.
- 25 Mandeep K. Dhami, Lars Wicke, and Dilek Onkal, "Scenario generation and scenario quality using the cone of plausibility," *Futures* 142, (2022): 7.
- 26 Dhami, Wicke, and Onkal, "Scenario Generation and Scenario Quality", 9.
- 27 Artner, Girven, and Bruce, "Assessing the Value", 1.
- 28 Artner, Girven, and Bruce, "Assessing the Value", 1.
- 29 Artner, Girven, and Bruce, "Assessing the Value", 1.
- 30 Artner, Girven, and Bruce, "Assessing the Value", 1.
- 31 Artner, Girven, and Bruce, "Assessing the Value", 4.
- 32 Mandeep K. Dhami, Ian K. Belton, and David R. Mandel, "The "analysis of competing hypotheses" in intelligence analysis," *Applied Cognitive Psychology* 33, no. 6 (2019): 1080.
- 33 Dhami, Belton and Mandel, "Analysis of Competing Hypotheses", 1085.
- 34 Dhami, Belton and Mandel, "Analysis of Competing Hypotheses", 1083, 1086.
- 35 United States Government, "A Tradecraft Primer", 1-40.
- 36 Coulthart, "Evidence-Based Evaluation," 376.
- 37 Coulthart, "Evidence-Based Evaluation," 379.
- 38 Coulthart, "Evidence-Based Evaluation," 379.
- 39 Coulthart, "Evidence-Based Evaluation," 375.
- 40 Chang et al., "Restructuring Structured Analytic", 338 - 339.
- 41 Chang et al., "Restructuring Structured Analytic", 338.
- 42 Incoherent probabilities is listed as the activated opposing bias for AFA, but not defined or expanded on further in the text, which focuses on base-rate neglect as the opposing bias of status-quo bias. Chang et al., "Restructuring Structured Analytic", 343-344.

- 
- 43 Chang et al., “Restructuring Structured Analytic”, 343-344.
- 44 Chang et al., “Restructuring Structured Analytic”, 340.
- 45 Chang et al., “Restructuring Structured Analytic”, 340.
- 46 Chang et al., “Restructuring Structured Analytic”, 344.
- 47 Artner, Girven, and Bruce, “Assessing the Value”, 5.
- 48 Stephen Marrin, “Evaluating the quality of intelligence analysis: By what (mis) measure?,” *Intelligence and National Security* 27, no. 6 (2012): 896.
- 49 Artner, Girven, and Bruce, “Assessing the Value”, 4.
- 50 United States Government, “A Tradecraft Primer”, 36.
- 51 Artner, Girven, and Bruce, “Assessing the Value”, 5.
- 52 Artner, Girven, and Bruce, “Assessing the Value”, 3, 12.
- 53 Artner, Girven, and Bruce, “Assessing the Value”, 12.
- 54 Marrin, “Evaluating the Quality”, 896.
- 55 Marrin, “Evaluating the Quality”, 908-911. Artner, Girven, and Bruce, “Assessing the Value”, 5.
- 56 Marrin, “Evaluating the Quality”, 897-899.
- 57 Richards J. Heuer Jr, “The evolution of structured analytic techniques,” (presentation to the National Academy of Science, National Research Council Committee on Behavioural and Social Science Research to Improve Intelligence Analysis for National Security, Washington, DC, December 8, 2009), 4-5.
- 58 Heuer, “Evolution of Structured Analytic Techniques”, 5.
- 59 Wing, “Short Introduction”, 20, 18.
- 60 Randolph H. Pherson, and Richards J. Heuer Jr, *Structured analytic techniques for intelligence analysis*, (Washington DC: Cq Press, 2011), 312.
- 61 Chang et al., “Restructuring Structured Analytic”, 345-346.
- 62 Wing, “Short Introduction”, 20.
- 63 Wing, “Short Introduction”, 21.
- 64 United States Government, “A Tradecraft Primer”, 36.
- 65 United States Government, “A Tradecraft Primer”, 36.
- 66 John Battersby, “Can Old Lessons Inform Current Directions: Australia, New Zealand, and Ananda Marga’s Trans-Tasman “Terrorism” 1975–1978,” *Studies in Conflict & Terrorism* 44, no. 8 (2021): 686-700.
- 67 Pherson, “Nine Techniques”, 5.
- 68 Chang et al., “Restructuring Structured Analytic”, 344.
- 69 Virginia Braun and Victoria Clarke, “Using thematic analysis in psychology,” *Qualitative research in psychology* 3, no. 2 (2006): 87; Virginia Braun and Victoria Clarke, “One size fits all? What counts as quality practice in (reflexive) thematic analysis?,” *Qualitative research in psychology* 18, no. 3 (2021): 328-352.
- 70 Braun and Clarke, “One size fits all?”, 333-334.
- 71 Artner, Girven, and Bruce, “Assessing the Value”, 14.
- 72 Artner, Girven, and Bruce, “Assessing the Value”, 1.
- 73 Chang et al., “Restructuring Structured Analytic”, 347, 345-346.
- 74 Chang et al., “Restructuring Structured Analytic”, 344.
- 75 United States Government, “A Tradecraft Primer”, 35.
- 76 Battersby, “Can Old Lessons”, 694.
- 77 Wing, “Short Introduction”, 20.
- 78 United States Government, “A Tradecraft Primer”, 35.
- 79 This response was received from a participant in round one, focus group five. This group participated in the focus groups, but their matrix was not recorded, and therefore could not be assessed. The group was still surveyed and their feedback on the process was collected.
- 80 Dhami, Wicke, and Onkal, “Scenario Generation and Scenario Quality”, 6-7.

- 81 David J. Brier, "Marking the future: a review of time horizons," *Futures* 37, no. 8 (2005): 833-848.
- 82 Dhami, Wicke, and Onkal, "Scenario Generation and Scenario Quality", 7.
- 83 Pherson, "Nine Techniques", 5.
- 84 Pherson, "Nine Techniques", 7,12.
- 85 Wing, "Short Introduction", 20.
- 86 Chang et al., "Restructuring Structured Analytic", 343-344.
- 87 United States Government, "A Tradecraft Primer", 36.
- 88 United States Government, "A Tradecraft Primer", 12.
- 89 United States Government, "A Tradecraft Primer", 12.
- 90 United States Government, "A Tradecraft Primer", 36.
- 91 United States Government, "A Tradecraft Primer", 34.
- 92 Chang et al., "Restructuring Structured Analytic", 344.
- 93 Hiltunen, "Was it a Wild Card", 62-63.
- 94 United States Government, "A Tradecraft Primer", 36.
- 95 Wing, "Short Introduction", 20-21.
- 96 Hiltunen, "Was it a Wild Card", 61.
- 97 Dhami, Belton and Mandel, "Analysis of Competing Hypotheses", 1085.
- 98 Chang et al., "Restructuring Structured Analytic", 345.
- 99 Heuer, "Evolution of Structured Analytic Techniques", 5.