Ascertainment of a hierarchy of dimensions from time-poor experts: Linking tactical vignettes to strategic scenarios

Brandon Pincombe a,⁎, Sarah Blunden b, Adrian Pincombe a, Patricia Dexter a

a Land Operations Division, Defence Science and Technology Organisation, PO Box 1500, Edinburgh, SA, 5111, Australia
b Appleton Institute, Central Queensland University, 44 Greenhill Road, Wayville, SA, 5034, Australia

A R T I C L E   I N F O

Article history:
Received 28 February 2012
Received in revised form 13 April 2012
Accepted 25 April 2012
Available online 27 May 2012

Keywords:
Delphi
Scenarios
Thematic analysis
Hierarchical dimensions
Mixed methods

A B S T R A C T

Many organisations define or express their strategic direction through strategic scenarios, which typically lack the detail for operational or tactical decisions. However such decisions need to be consistent with the organisation’s overall strategic direction. We present a method for defining a dimensional hierarchy of important variables at the strategic, operational and tactical levels and illustrate this method with a case study. Achieving a set of dimensions as comprehensive as possible requires the saturation of ideas from all relevant expert groups. Unfortunately, access to experts may be limited and sporadic. We facilitate expert participation through an online, iterative, Delphi-like structured group technique allowing asynchronous access from many separate locations. Inter-round thematic analysis removes duplicate argumentation, standardises language conventions and builds a hierarchy of important variables: strategic dimensions; operational components; and tactical elements. Iteration allows experts to assess and alter these themes. Anonymity and thematic de-duplication allow merit-based discussion by masking the status and numbers of advocates. In our case study, sixty-nine experts developed a hierarchical dimensionalisation of future Australian Army deployment contexts over three rounds with a total median time commitment of less than 3 h per expert.

1. Introduction

Preparatory scenarios are sketchy representations of plausible futures that can be used to evaluate alternative strategies for achieving desired objectives or develop sensitivity to the precursors of events requiring contingency actions [1]. They differ from predictive scenarios in that no one expects them to eventuate but are disturbingly similar in that both only have an infinitesimal chance of actually occurring. The power of preparatory scenarios comes from their representation of future uncertainties. Most scenario method practitioners choose to represent this uncertainty with a small number of scenarios built to frame the extreme outcomes of the ‘dimensions’, ‘key factors’, ‘critical uncertainties’, ‘driving forces’, ‘causal factors’ or ‘important variables’ [1-8]. Others use a single super complex scenario simultaneously containing the extreme dimensional outcomes [9,10]. However, there is general agreement that an explicit dimensionalisation, performed by experts, is an important early step in scenario thinking [1-7,9,11-16]. Correctly choosing the key uncertainties is critically important as strategies that are robust, resilient, redundant and flexible across one set of uncertainties will not necessarily be so across a different set of uncertainties. The quality and representativeness of experts is critical in the selection of the right dimensions [1], as is the case for scenario [17-19] and Delphi [20,21] methods in general. Good scenario methods consult experts in a manner that supports diversity and seek the most general dimensions by using a two stage process that clusters or groups the dimensions suggested in the first stage [1]. As experts have limited and sporadic availability [9], the best possible use must be made of their limited time.

⁎ Corresponding author.
E-mail address: brandon.pincombe@dsto.defence.gov.au (B. Pincombe).

0040-1625/$ – see front matter. Crown Copyright © 2012 Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.techfore.2012.05.001
Scenario thinking does not stop at the strategic level. More focused scenarios, that should be consistent with the organisation’s overall strategic direction, are often used at lower organisational levels. Names for hierarchical levels vary across domains and, confusingly, commonly used hierarchies include strategic–tactical–operational and strategic–operational–tactical. We use the latter because of our defence backgrounds and address the problem of consistency. Our method of dimensional development allows involvement from many people managing lower levels of the organisation and hierarchically groups the driving forces that they identify as strategic dimensions, operational components and tactical elements. This provides a greater diversity of ideas and points of view in the development of strategic dimensions but it also gives a starting point for developers of operational scenarios and tactical vignettes that is consistent with the organisation’s strategic dimensions.

Delphi based methods for scenario building, both of the Ariadne and on-line [22] and online [23,24], have, thus far, presented experts with dimensions pre-defined by the researchers, moderators or analysts. An online Delphi method by which experts could nominate dimensions would extend the credibility of existing online Delphi methods for scenario building by involving experts from the very start of the scenario construction process.

We present an online Delphi process for dimensionalising problem spaces that could be used as a first step in online scenario development and allows the construction of hierarchical dimensional frameworks. This process is an iterative structured group technique incorporating thematic analysis [25] between rounds, to standardise language conventions and remove duplicate argumentation, that allows a consensus of dimensions, components and elements to be achieved by experts working asynchronously at separate locations. Experts are either fully engaged or free for other work. Anonymity allows fearless presentation of ideas but also masks the status of experts allowing the ideas of less dominant people to be considered on their merits, unlike workshops where status hierarchies rapidly form even if they do not already exist [26]. Arguments are themed [25] by moderators between rounds and a single consolidated statement of each theme is presented in the subsequent round without any indication of dominance, confidence or majority. Not only does this allow panelists to use their expertise to evaluate arguments on their merits [27] but de-duplication also saves experts time by eliminating repetition. The iterative nature of the technique allows the participants to correct themes that are poorly expressed or entirely missed by the moderators.

Ascertaining dimensions from experts is a qualitative process and its trustworthiness is established through credibility, dependability, confirmability, applicability, ecological validity and ethical behaviour [21]. Credibility [21] or ‘internal validity’ [21,28] is demonstrated through panelists quantitative confirmation of moderators interpretations of their qualitative responses from the previous round [21,28] and through methods to ensure merit-based assessment of argumentation [27,29]. Dependability is based on selection of participants to include a representative sampling of expertise that may affect the outcome [17-21,30] with the motivation to contribute [18,30,31]. Confirmability [21] or auditability [28,32-35] entails the recording of a detailed and honest description of the collection and analysis process, even if this process diverges from the protocol. Applicability [28], sometimes called transferability [21,36] or ‘external validity’ [37], is the extent to which the findings make sense in a broader context and agree with those derived from other methods such as interviews [38,39], focus groups [40,41], nominal groups [42], further questionnaires beyond the Delphi participants [43] or narrative analysis [37]. Ecological validity requires that questions be asked in a manner, language and setting similar to that participants are used to in the everyday exercise of their expertise [19,44]. The appearance [31] and presence of ethical behaviour, as evidenced by treatment of participants and ethics committee endorsement, is also critical to a trustworthy process. All of this presupposes that expertise is indeed learnable in the area under consideration [19]. Our Delphi process is aimed at producing a hierarchical dimensionalisation of a given problem-space. Scenario developers must remember that it is only an indication of the opinions of those experts consulted and so cannot replace the need for judgement on their part [45].

Later in this paper we demonstrate this process through a case study stepping through the first stages of a process to find the strategic dimensions, operational components of these dimensions and tactical elements of these components for the contexts in which the Australian Army may be deployed. Three rounds were needed to generate consensus dimensions. At the end of three rounds there was consensus on thirty-four operational components of the dimensions but consensus on whether this was a complete set was not asked. A low time commitment of approximately an hour per expert per round was demonstrated.

2. Method

We present an online method for determining a dimensional hierarchy linking important variables at the strategic, operational and tactical levels of an organisation. The form of this linkage is participant driven and is hierarchical only in the sense that operational problems are more detailed and focused than strategic ones. This method can be seen as a conflation and expansion of several early steps in standard scenario thinking methodologies, but does not explicitly include the earliest step of defining the scope [4] or setting the scenario agenda [1]. It presupposes that researchers and sponsors have arrived at an agreed focal issue and time frame. Our method is focused on developing a structured hierarchy of strategic dimensions, operational components and tactical elements to improve the consistency of scenarios at the various levels and make it clear where these fit within an overarching framework. However, minor modifications, especially progressive removal of the operational and tactical details, would allow it to identify major stakeholders, basic trends and key uncertainties and construct initial scenario themes (steps two to five in Schoemaker’s ten-step scenario method [4]). Similar changes would allow it to determine and cluster the driving forces and a small addition, detailed later, would allow definition of the cluster outcomes (steps two to four of Wright and Cairn’s eight-step scenario method [1]). As it stands, the method iterates between online questioning of participants and thematic analysis by researchers to arrive at an agreed set of themes, sub-themes and elements while minimising time wasting repetition of these to participants.
2.1. Drafting the first round questionnaire

Initially a draft questionnaire needs to be constructed for the first round of the Delphi process. The nature of this questionnaire has a considerable bearing on the success of the process [30]. Most Delphi theorists advise moderators to start with open ended qualitative questions rather than a prestructured set of options as the latter has the potential to severely circumscribe the set of possibilities considered by the panel [30]. Hill [30] concentrates on the inability of experts to raise options beyond those contained in the pre-structured set if Delphi begins with an entirely quantitative survey. Tversky and Kahneman’s anchoring-and-adjustment heuristic [46] raises the possibility that the suggestion of new options by experts may be corrupted by the presentation of prestructured options in the first round of the survey. However, the dimensions of a problem-space are more complex constructs than the point estimates considered in most anchoring literature. We contend that experts that already have a strong, but potentially implicit, mental model of the dimensions of a problem-space are unlikely to be swayed by anchoring. The ability of experts to avoid anchoring is supported by the behaviour of experts in the case study that we report below. Therefore, we suggest a first round survey in which the dimensions mentioned in the literature are presented for the experts to assess on a 5-point Likert-type scale and make qualitative comments on but that also contains a section asking for new suggestions for dimensions beyond those in the prestructured set and a section asking for the four to six top level dimensions that the expert would use to characterise the scenario-space.

We have found that experts, even those with a good understanding of the use of scenarios, often do not know what is being looked for when moderators ask them to nominate dimensions. Dimensions need to be described to the experts but, just as we use a case study in this paper to further explain this method, examples of dimensions will aid experts’ understanding of the concept. A set of prestructured questions nominating the likely dimensions based on the literature in the area gives participants a chance to see by example what is being asked for. Unfortunately, Tversky and Kahneman’s anchoring-and-adjustment heuristic [46] raises the possibility that by suggesting dimensions the experts will anchor on these dimensions and be less likely to suggest those that they would independently construct as a mental model of the scenario-space. Expertise has been shown to reduce, but not eliminate, the effects of anchoring in point financial market predictions [47] and participants with low expertise are disproportionately affected by anchoring in both point and range prediction questions about Australian Rules Football [48]. The degree of anchoring and adjustment in complex constructs, such as dimensions, is difficult to test. We contend that there will be little anchoring as long as the experts have truly learnt judgement in their area of expertise and are motivated to share this expertise. Therefore, we propose that an expert advisory panel should iteratively develop questions presenting candidate dimensions mentioned in the literature and allowing participants to assess their relevance and importance on 5-point Likert-type scales as well as providing an opportunity to provide qualitative comments on each candidate dimension. True experts are likely to have read or heard about these proposed dimensions and have pre-formed opinions on them. In the unlikely chance that the set of dimensions presented is excellent, or the experts do anchor strongly, the process can end with the first round.

After giving a definition of a “dimension” and presenting as complete as possible list of candidate dimensions from the literature as examples, experts should be asked for their dimensionalisation of the problem-space. As anchoring is a potential problem, we advise moderators to ask the participants to nominate and describe new candidate dimensions immediately after the prestructured options. Then the participants can be asked to name and describe the four to six top level dimensions that they would use to characterise the problem-space.

To improve the credibility and dependability of the draft first round questionnaire it should be iterated through an expert advisory panel made up of the moderators [49] as well as an expert [49] and a representative of the sponsor organisation that requires dimensions for their scenarios [9]. This process should be recorded for auditability. Contributions by the expert and sponsor are intended, amongst other things, to raise the ecological validity [19] of the final draft of the questionnaire. The appropriate time to formalise the protocol and submit it to the relevant ethics committee is after the expert committee has completed the draft of the first round questionnaire. Not only will the ethics committee act as a further credibility check on the content of the questionnaire but also contains a section asking for new suggestions for dimensions beyond those in the prestructured set and a section asking for the four to six top level dimensions that the expert would use to characterise the scenario-space.

After giving a definition of a “dimension” and presenting as complete as possible list of candidate dimensions from the literature as examples, experts should be asked for their dimensionalisation of the problem-space. As anchoring is a potential problem, we advise moderators to ask the participants to nominate and describe new candidate dimensions immediately after the prestructured options. Then the participants can be asked to name and describe the four to six top level dimensions that they would use to characterise the problem-space.

To improve the credibility and dependability of the draft first round questionnaire it should be iterated through an expert advisory panel made up of the moderators [49] as well as an expert [49] and a representative of the sponsor organisation that requires dimensions for their scenarios [9]. This process should be recorded for auditability. Contributions by the expert and sponsor are intended, amongst other things, to raise the ecological validity [19] of the final draft of the questionnaire. The appropriate time to formalise the protocol and submit it to the relevant ethics committee is after the expert committee has completed the draft of the first round questionnaire. Not only will the ethics committee act as a further credibility check on the content of the questionnaire but approval is also an external indicator that the research is ethical. This is important as it provides an external guarantee that the changes made by the moderators are the changes asked for by the content validity experts. Therefore, we propose that an expert advisory panel should iteratively develop questions presenting candidate dimensions mentioned in the literature and allowing participants to assess their relevance and importance on 5-point Likert-type scales as well as providing an opportunity to provide qualitative comments on each candidate dimension. True experts are likely to have read or heard about these proposed dimensions and have pre-formed opinions on them. In the unlikely chance that the set of dimensions presented is excellent, or the experts do anchor strongly, the process can end with the first round.

After giving a definition of a “dimension” and presenting as complete as possible list of candidate dimensions from the literature as examples, experts should be asked for their dimensionalisation of the problem-space. As anchoring is a potential problem, we advise moderators to ask the participants to nominate and describe new candidate dimensions immediately after the prestructured options. Then the participants can be asked to name and describe the four to six top level dimensions that they would use to characterise the problem-space.

2.2. Piloting the questionnaire and delivery mechanism

The ecological validity of the survey and the learnability of the domain determine the maximum possible level of expert performance [19]. To ensure the draft questionnaire is written in language that is understood by the experts who will participate in the survey, a small group of similar experts should be selected to undertake a content validity process [49]. Content validity can be assessed, and ecological validity improved, by presenting each of the questions in the questionnaire to the experts with an opportunity to vote “yes” or “no” on whether it should be included in the final survey and a chance to comment on how they feel the question should be changed [49]. Anonymity of participation reinforces the credibility of this process because arguments are heard based on their merit rather than on the status of those making them [27]. Credibility may be able to be improved by iterating this validity process to ensure that the changes made by the moderators are the changes asked for by the content validity experts. However, the expert representation on the committee developing the draft of the first round questionnaire should mean that most survey questions will be phrased in an ecologically valid manner. The closer the background and experience of the validity experts to the expert participants in the main round, the higher the dependability of this process [49]. The content validity assessment process also gives an opportunity to test the online delivery tool chosen for the main study in a situation where less participants are
2.6. How many Delphi iterations?

Delphi practitioners choose many different conditions to indicate when to cease iterating through Delphi rounds [29,31,37,45,53,54]. Most Delphi processes have a qualitative stopping condition: consensus is considered to have occurred when moderators assess expert responses to have changed little between rounds [54-56]. Williams and Webb’s survey of consensus conditions [54] describes a quantitative definition based on the standard deviations in Likert-type scores, but even this uses “the data as an empirical guide for the level of consensus rather than defining the standard prior to study” [54]. Thus consensus is arbitrarily decided by the researchers, and ecological validity that and the online delivery tool works. Once this has been determined, the main study can commence.

2.3. Recruiting and retaining experts: maximising their convenience

The truthfulness of dimensions depends on the availability of a representative sample [17-21,30] of experts who have learnt judgement in the problem-space [19] and are motivated to contribute [18,30,31]. Access to experts may be both limited and sporadic [9]. Our Delphi approach facilitates expert participation by its online nature, allowing asynchronous access from many separate locations. The use of thematic analysis to de-duplicate responses between rounds does not simply increase credibility through removing cues as to numerical support for a particular position [27]. It also reduces the time commitment from experts as they only need to listen to each argument or consider each theme once. The sponsors of the study, who desire a dimensionalisation of their problem-space, are typically the best people to aid researchers in accessing a representative sample of experts. Once the participants have been selected and approached, the first round Delphi survey can be delivered to those who give informed consent to participate.

2.4. Thematic analysis: constructing the dimensional hierarchy

Constructing a dimensional hierarchy is a qualitative process. Thematic analysis [25,50,51] captures important aspects of qualitative data sets without being wedded to an existing theoretical framework [50]. Because of this independence, we suggest that it be used in the analysis of first round responses to derive a representation of the dimensions, components and elements that participants believe should be used to construct scenarios. While online implementations of Delphi are able to conduct a continuous interaction between experts [52], this eliminates the capacity for researchers to theme expert responses between rounds. The need to perform thematic analysis to structure the qualitative contributions of experts is the reason why we use a traditional round-based Delphi process.

Delphi participants communicate in writing so there is no need for a time consuming and error prone transcription process. Thematic analysis, therefore, begins with the data-driven generation of the initial codes identifying the most basic element of the text that provides meaningful information about the dimensional structure of the problem-space [25]. After the codes are developed, they may be grouped [25,50] into themes and sub-themes. The expert participants are likely to give strong cues as to how such groupings should be constructed. This process yields a hierarchy of themes, sub-themes and elements.

The dependability and credibility of this thematic hierarchy is crucial to a trustworthy Delphi process. Researchers with expertise in the problem-space are likely to be more dependable because they have the experience to recognise appropriate codes in the textual responses and sufficient acquaintance with problem-space specific vocabulary to phrase codes in an ecologically valid manner. The credibility of the coding process can be tested by comparing the codes, sub-themes and themes developed by multiple coders from the same text. The Delphi method permits large numbers of participants, so there may be too many texts to all be coded by all coders. Inter-rater consistency can be tested by randomly selecting a small set of responses for all researchers to code and theme before the rest of the responses are considered. If there is high inter-rater consistency of code and theme development on this sub-set the remainder of the textual responses can be coded by a subset of researchers. To allow continued credibility checking the researchers should still be paired so that at least two of them code and theme each text. The credibility of the codes can be enhanced still further if saturation of codes and themes can be shown [25,49] through the tailing off to zero of the rate of appearance of new codes or themes as more texts are considered. Saturation is good for credibility as it means that further responses from a larger group of similar experts are unlikely to contain more themes.

Following thematic analysis the researchers should have a tentative statement of the dimensions, components and elements of the problem-space. The elements identified through coding represent a candidate set of elements relevant to constructing tactical vignettes. Those sub-themes identified represent a candidate set of components relevant to the construction of operational scenarios. The key themes form a candidate list of dimensions for strategic scenarios. When this hierarchical dimensional construct is represented to the Delphi panelists for their criticism, they have no idea of the number, status or dominance of the proponents of each code, sub-theme and theme so that they can be assessed on their merit alone [27].

2.5. Dimensional importance is variable across levels

Consensus may be reached and the Delphi process completed at any round after a dimensional hierarchy is constructed and put to the experts for a quantitative vote. In any round that may plausibly be the final round, participants should be asked to rank the importance of each of the dimensions at the tactical, operational and strategic levels. This provides extra information to scenario or vignette builders on which dimensions, components or elements are most critical to include.

2.6. How many Delphi iterations?

Delphi practitioners choose many different conditions to indicate when to cease iterating through Delphi rounds [29,31,37,45,53,54].
introducing subjectivity and reducing credibility [54]. Furthermore, researchers often only report the degree of consensus in descriptive terms such as ‘high’ rather than reporting the actual figures so that readers can make their own decisions about whether consensus is sufficient [54]. We resolve to decide on a degree of agreement needed to define consensus as the first round survey is distributed and to report the actual values of agreement.

“Consensus” rates of between 50% and 100% are commonly used in Delphi studies [54] and rates between 80% and 100% are widely seen as being acceptable in most cases [54]. We propose a sliding scale of consensus based on the number of participants expected to respond in the final round. When thirty or more participants are expected to remain in the final round we suggest that consensus be defined as ≥90% of participants concurring to a statement. If fifteen or more, but less than thirty, participants are expected to remain in the final round we suggest that three be able to dissent with a statement and for it still to be considered to have achieved consensus support. If less than fifteen participants are expected to be retained to the final round we suggest that more participants be recruited. Dissent is defined as “no” or “unsure” in a yes-no scale or “Neither Agree nor Disagree”, “Disagree” or “Strongly Disagree” on a 5-point Likert-type scale and concurrence as the complement of dissent. Setting the consensus level at the start of the first round requires an estimate of the drop-out rate over the process. Delphi is renowned for its high attrition rate [57] so we assume that only one third of participants starting the first round will continue to participate throughout the process [57]. Thus, if there are between 45 and 89 first round participants, the consensus rate is determined by dividing the number of first round participants by three and working out a percentage allowing for three dissenters. Therefore, the consensus rate of concurrence varies from 80% to 90% in a manner defined as the study commences.

The first round of our process bears similarities to a Policy Delphi [58], aiming to maximise the range of expert relevant opinions garnered [59,60], which may seem like dissensus. Thematic analysis allows these opinions to be re-presented to participants on an equal footing regardless of their support so that they can be assessed on their merit alone [27]. However, if the researchers and sponsors have clearly defined the focal issue then the participants should iterate towards a group consensus on the driving forces in the contextual environment of the sponsors’ organisation for that focal issue. Continuing dissensus or the formation of cliques is either a sign of a poorly defined focal issue or of an issue for which experts cannot agree on what the critical uncertainties are. In the first case the statement of the focal issue should be revised with the sponsor and clarified. The second is the equivalent of a total inability to identify key uncertainties in Schoemaker’s scenario method [4] or cluster the driving forces in Wright and Cairn’s scenario method [1] and may indicate too divisive a problem for scenario thinking.

2.7. Assessment of applicability

Applicability [28], transferability [21,36] or ‘external validity’ [37], is the extent to which the findings make sense in their broader context. This can be determined by agreement with findings derived from other methods such as interviews [38,39], focus groups [40,41], nominal groups [42], further questionnaires beyond the Delphi participants [43] or narrative analysis [37]. We suggest that, in the first instance, findings be passed back through the expert panel established to draft the first round questionnaire for them to make a preliminary assessment of applicability.

3. Case study: Australian Army deployment context dimensions

The Australian Defence Force (ADF) uses three primary Australian Contingency Context Scenarios (ACCS), representing the essential features of contexts present in Australia and its ‘inner-arc’ of neighbours it might aid, to test all proposed changes to the structure of the force or the capabilities it employs. Tied options are separated using two secondary ACCS, based on those deployment contexts only found outside Australia’s immediate region. A committee of senior scenario analysis experts drawn from across the ADF develop, and review, these strategic scenarios. No formal definition of the dimensions of the scenario-space was undertaken because the committee members, steeped in the same doctrine and working together for many years, had a shared, but implicit, understanding of these dimensions. However, force structure and capability proposals are best tested at the operational or even tactical levels and ADF analysts spend considerable effort developing operational scenarios and tactical vignettes that are consistent with the strategic-level ACCS. A hierarchical dimensional structure would provide a framework for this process.

We worked with the Director General of Future Land Warfare (DGFLW) to trial our process on the Australian Army and develop a hierarchical dimensional framework that, at its highest level, informs reviews of the ACCS but that also provides guidance on how to remain consistent with strategic intent in operational scenarios and tactical vignettes. As the ACCS are joint scenarios across the Navy, Army and Air Force, this process needs to be repeated with joint participants to produce an appropriate hierarchical dimensional framework for the ACCS rather than information on the specific requirements of the Army.

3.1. Drafting the first round questionnaire

The lead author formed a draft survey presenting examples of dimensions from similar problem-spaces [61-67] and asking participants to construct dimensions for future land warfare contexts for the Australian Army. An expert panel, composed of the DSTO authors, the senior military officer at DSTO and our Future Land Warfare (FLW) desk officer, twice reviewed and revised this questionnaire before it was written into a protocol submitted to the Science and Technology Activity Review (STAR) sub-committee of the Australian Defence Health Research Ethics Committee (ADHREC). Ethics approval occurred after three iterations through STAR in which a number of demographic questions that may have identified minority participants were removed and participation was limited to ADF members. This produced a draft questionnaire, ready for a validity study, containing eight
demographic questions, thirty-one statements describing example dimensions, two free response questions asking for ideal and extra dimensions, a question on agreement with FLW’s categorization of the roles of the Australian Army and four questions about interactions with other agencies.

3.2. Piloting the questionnaire and delivery mechanism

The draft survey was modified in accordance with the results of a validity questionnaire in which six consenting experts were asked whether questions should be retained, wording changed and new questions added. Participant rank was recorded on consent (see Table 2) but no demographic information was asked due to STAR concerns about identification risks. Five participants answered the online questionnaire served using Survey Monkey. Two returned printed surveys, one of which indicated it contained further comments that could not be entered into Survey Monkey as the participant had finalised submission and the other being very brief. Survey Monkey was shown to work as a delivery mechanism on the ADF intranet. After taking into account the votes and comments of participants a final form of the first round survey was constructed. The demographic questions in this survey now directly asked for years of military experience and places deployed rather than presenting a drop-down list. Main study participants were to be asked to rate agreement with the statements “this is an enduring dimension of future land warfare for the Australian Army” and “this is important to consider in future land warfare scenarios for the Australian Army” on a 5-point Likert-type scale (“Strongly Agree”, “Agree”, “Neither Agree nor Disagree”, “Disagree” or “Strongly Disagree”) for each of the 34 proposed dimensions as well as to make free response comments on them. Two free response questions asked for names and sentence length descriptions of “up to four dimensions of future land warfare for the ADF” that had been missed in the preceding list and “the six most important dimensions to use when building scenarios for future land warfare involving the Australian Army”. Finally, the survey asked for a ‘yes’ or ‘no’ answer and a comment on whether “the following five categories cover the roles that are expected to be performed by the Australian Army: conventional war; counter insurgency; stabilisation; peace support; and humanitarian” as well as alternate categorisations.

3.3. Main study

Participants undertook three rounds of surveys to iteratively refine a group view of the dimensions of the contexts in which they expect the Australian Army may be required to deploy soldiers into in the future. They could respond when and where they wanted and participate for as long as they felt they had more to say while the survey was open. Researchers constructed the second and third questionnaires in the eleven days between rounds based on their analyses of the qualitative and quantitative responses from the previous round. Saturation, defined as the appearance of no new themes [25], was looked for in the analysis of each round. At the end of the third round not only did researchers fail to detect new themes but also no participant opposed the statement that “no important dimensions were missing”.

3.3.1. Participants

Enumerating the critical uncertainties that must be considered to prepare robustly for the future employment of any army is a difficult task requiring deep expertise in the factors effecting combatants as well as those supporting and supplying them and repairing their equipment. It takes years of study and experience to learn good judgement in each of the important sub-elements of this domain: sub-elements best defined by reference to the corps structure of armies. On our behalf, DGFLW approached 93 experts and 25 postgraduate students to participate. Experts included all members of Force Development Group (FDG) and the members of FLW with a similar role to FDG. These experienced officers, drawn from across the corps of the Australian Army in a representative manner, have the task of developing future force structures and testing them using scenarios. The remainder of the experts were the authors of future focused papers in the previous three volumes of The Army Journal. All 25 Australian Army officers undertaking a masters degree at the Australian Command and Staff College were also approached to participate in the first round, which was timed so they had just completed a fortnight-long block-release course in military futurology involving 60 h of lectures and tutorials and were undertaking a week of participation in military experimentation to further familiarise themselves with the use of scenarios. Informed consent was given by 73 experts and 22 students who were free to withdraw from the study or cease participation without prejudice. Of these 22 experts and 4 students ceased participation before starting the first round. The anonymous survey design meant that those who ceased participation could not be identified, even by the researchers with full data access. All students dropped out after the first round. We worked on the expectation that only one-third of the 69 first round participants would remain until the end of the survey and set a consensus level such that concurrence (“Strongly Agree” or “Agree”) would be required from all but 3 of the 23 participants expected to remain, that is a consensus level of 87%. Students and experts were experienced (see Table 1) regular officers in the Australian Army. Expert participants were high ranking (see Table 2). Given the manner in which members of FDG are selected it is unsurprising that all combat and combat support corps were represented (see Table 3) as were most combat service support corps. This diversity indicates deep expertise in the participant pool across the entire battlespace.

3.3.2. Conduct of the first round

Participants were sent a group e-mail by the study sponsor, explaining the reasons for the study and exhorting them to participate. Personalised e-mails followed containing information and consent forms with return information. Informed consent was given by 22 of the 25 students and 73 of the 93 experts. Three consent forms were faxed to the (prominently displayed) ADHREC complaints number, but ADHREC follow-up determined this was by mistake and they were passed on to the principal researcher. Consenting
participants were sent individual e-mails containing another copy of the information sheet; extra reminders that they were free to drop out at any time and that only their responses, not their participation, was anonymous; instructions to follow their local OH&S rules; a URL leading to a Survey Monkey collector; and a printable MSWord version of the survey with instructions on where to post it. Different collectors were sent to the student and expert groups. Participants were given two weeks and two days in which to respond. Five days before the end, personalised reminder e-mails were sent to all participants. These again contained the appropriate URL but also had an apology to those who had already completed the survey pointing out that anonymity meant the researchers did not know who had. No consenting participants dropped out. No printed surveys were returned. One respondent reported problems re-accessing Survey Monkey after taking a break. Of the consenting participants, 18 students and 51 experts commenced the survey and 18 students and 50 experts completed the entire survey. Excluding those who left the survey window open for at least one entire night the time spent with the survey window open ranged from 3 minutes to 295 minutes with a median of 32 minutes for the students and from 2 minutes to 361 minutes with a median of 55.5 minutes for the experts.

### 3.3.3. Analysis of the first round

Only 3 of the 34 candidate dimensions achieved the greater than 87% level of support (“Agree” or “Strongly Agree”) we use as a consensus level. We examined the qualitative responses to see why. The primary reason was that the prospective dimensions, while often descriptive of future land warfare contexts for the Australian Army at the operational or tactical levels, were too detailed to be strategic dimensions. We applied thematic analysis [25,50,51], as in [49], to the qualitative responses contained in the participants comments on the thirty-four candidate dimensions, their paragraph long descriptions of extra dimensions that they thought had been missed out and their paragraph long descriptions of the one to six dimensions that they would use to describe the important features of future land warfare contexts for the Australian Army.

Qualitative responses were allocated to five sets for thematic analysis by the DSTO authors and another DSTO member. Responses of individuals were kept together and the five sets were formed by random allocation of participants using pseudo-random numbers on the [0,1] interval [68]. The sets were defined as $b_{0.2}$, $b_{0.4}$, $b_{0.6}$, $b_{0.8}$ and $b_{1.0}$ with responses allocated to them until they were full (four students or eleven experts) at which point they were removed from further allocations. The first of these sets, with four student and ten expert responses, was independently themed by all four researchers. Disagreements between raters were limited to the wording of theme descriptions and the level of detail. Therefore, the other four sets—termed sets A, B, C and D—were each themed by two researchers: one researcher themed sets A and B; another B and C; another C and D; and the last D and A. Saturation was not achieved as there were three themes (access to purchased goods and services [i.e. blockade, sanctions or rationing], economic cycles and informality of the economy) that were suggested by only one respondent. The themes derived from these two questions are shown in Table 4 as components of newly formed higher level dimensions. The student group suggested a subset of the themes suggested by the expert group, consistent with the student group being both smaller and less experienced.

The newly developed themes were combined with the original themes and were grouped by the researchers into sets that defined the new candidate dimensions. This grouping was informed by the textual responses of the participants both in the two exclusively qualitative questions and in the comments to the initial set of 34 questions. Space limitations prevent the full descriptions of the dimensions being given but their names are given in the first four columns of Table 4 and are indicative of the level of agreement arrived at. The names of the final dimensions are given in bold in the fifth column of Table 4 and the names of their constituent components follow in square brackets. Again, space limitations prevent a detailed description of the components.

The final question, asking whether participants agreed that the roles covered by the Australian Army are conventional war, counter insurgency, stabilisation, peace support, and humanitarian operations gathered 31 ‘yes’, 25 ‘no’ and 13 ‘no response’ answers. The qualitative results indicated a diversity of reasons for voting ‘no’ or for being less than enthusiastic about having voted ‘yes’. These include: a philosophical opposition to categorisation of warfare into types; a view that the Army will need to do several (and possibly all) types of operation at the same time; a different categorisation scheme, particularly one based on the five lines of operation (see [67]) – joint land combat, population protection, population support, indigenous capacity building and information actions; or a more

### Table 1

<table>
<thead>
<tr>
<th>Years as a member of the military.</th>
<th>10 or less</th>
<th>11 to 14</th>
<th>15 to 19</th>
<th>20 to 23</th>
<th>24 to 27</th>
<th>28 or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experts</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>19</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Students</td>
<td>0</td>
<td>2</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Rank</th>
<th>Validity</th>
<th>Round one</th>
<th>Round two</th>
<th>Round three</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experts</td>
<td>Students</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colonel</td>
<td>1</td>
<td>10</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Lieutenant Colonel</td>
<td>0</td>
<td>31</td>
<td>26</td>
<td>17</td>
</tr>
<tr>
<td>Major</td>
<td>4</td>
<td>9</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Civilian</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No answer</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
extensive categorisation scheme, especially including domestic security and chemical, biological, radiological and nuclear (CBRN) weapons issues.

To assess the similarity between the 18 students and 51 experts, the responses to the 68 Likert-type items were combined into Table 5 and a $\chi^2$ test [69] was performed. To test whether students and experts were drawn from different populations we established the null (H0) and alternate (H1) hypotheses:

**H0.** The students and experts are drawn from the same population and

**H1.** The students and experts are drawn from different populations.

With four degrees of freedom, the $\chi^2 = 21.903$ calculated from Table 5 gives $p < 0.001$ for rejecting $H_0$ and accepting $H_1$. Thus we can say with considerable certainty that students and experts are drawn from different populations. The major contributors to the high $\chi^2$ test score were the “Disagree” and “Neither agree nor disagree” indicating that the students were considerably more agreeable than the experts. This is consistent with the 493 challenging comments from the experts (9⅔ per-expert) compared to 55 for the students (3 per-student). The Cronbach’s Alpha [70,71] on these Likert-type items was $\alpha = 0.960$, which implies very little item specific variance and thus a high reliability.

### Table 3
Corps membership of first round participants (two staff officers also indicated their former corps).

<table>
<thead>
<tr>
<th>Corps</th>
<th>Experts</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staff</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Royal Australian Armoured Corps</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Royal Australian Artillery</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Royal Australian Engineers</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Royal Australian Corps of Signals</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Royal Australian Infantry Corps</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Australian Army Aviation</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Australian Army Intelligence Corps</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Royal Australian Corps of Transport</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Royal Australian Army Nursing Corps</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Royal Australian Army Ordnance Corps</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Royal Australian Electrical &amp; Mechanical Engineers</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Royal Australian Corps of Military Police</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Australian Army Public Relations Service</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>No Answer</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 4
Columns one to four show the thematic names independently selected by the four analysts. The fifth column shows the analysts consensus on the names of the dimensions (themes) in bold and components (sub-themes) in italics.

<table>
<thead>
<tr>
<th>Analyst</th>
<th>Dimensions: components</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME AP PD BP</td>
<td></td>
</tr>
<tr>
<td>Physical environment</td>
<td>Physical environment; Climate/Weather; Strategic distance; Lines of communication within the area of operations; and infrastructure</td>
</tr>
<tr>
<td>Human</td>
<td>Complex human terrain: Density of non-combatants; Population allegiances; Heterogeneity; Information Environment; Language skills and barriers; Civil security; and Civilian neediness</td>
</tr>
<tr>
<td>Operational partnerships</td>
<td>Operational partnerships: Other agencies; Interoperability; Partner support; Contractors; and Host Nation Capacity</td>
</tr>
<tr>
<td>Political Environment</td>
<td>Politics and Alliances: Host nation governance; Enemy politics; Access to purchased goods and services; National interest; Support and legitimacy; Endstates; Legal issues; Media access; and Loss of alliance</td>
</tr>
<tr>
<td>Economics</td>
<td>Economics: Economic system; Informality; Funding sources; Resource issues; and Cycles</td>
</tr>
<tr>
<td>Threat</td>
<td>Nature, diversity, intensity of the threat</td>
</tr>
<tr>
<td>Limits and capabilities of forces</td>
<td>Limits, capabilities and options for all forces</td>
</tr>
</tbody>
</table>
Table 5
Consolidated agreement levels for students and experts over 68 Likert-type questions.

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>3</td>
<td>61</td>
<td>105</td>
<td>535</td>
<td>384</td>
</tr>
<tr>
<td>Experts</td>
<td>17</td>
<td>246</td>
<td>387</td>
<td>1356</td>
<td>936</td>
</tr>
</tbody>
</table>

3.3.4. Design of the second round

The responses to the first round allowed construction of a ten page online second round survey. The first page asked for the rank of the participant. The next seven pages each dealt with a single dimension from Table 4. Detailed descriptions of all proposed components were presented for each dimension. After the description, participants were asked for a “yes”, “no” or “unsure” response to the two statements “do you agree that the concept expressed in this component should be included in this dimension” and “do you agree with the way this component is worded” as well as for their comments and suggestions. At the end of the page were two further questions. The first asked for a “yes”, “no” or “unsure” response to the two statements “are we missing something” and “should this be split” as well as asking for comments or alternate solutions. The second presented statements that summarised controversial or important issues raised in textual responses to the first round so that all participants could indicate their agreement on a 5-point Likert scale. The ninth page gave an opportunity to list extra dimensions or describe alternate dimensionalisations and the tenth asked for final comments.

3.3.5. Conduct of the second round

The second round commenced eleven days after the first round closed. Personalised e-mails were sent to each of the 95 consenting participants containing a URL leading to a Survey Monkey collector and indicating that they had two weeks and two days to complete the survey and were free to drop out or simply not respond. Collectors were constructed for both the student and expert groups even though the student group was expected to drop out at this stage. Individual reminders containing links to the appropriate collector were sent to experts six days before the end and to students five days before. No one dropped out but no students responded. The rank details of the thirty-nine expert respondents are recorded in Table 2. Excluding those who left the survey window open for at least one entire night the time spent with the survey window open ranged from 1 min to 370 min with a median of 65 min.

3.3.6. Analysis of the second round

The qualitative elements of all thirty-nine responses were analysed by the four authors independently. The statements following each component were considered in conjunction with the support for the retention of that component, the support for the retention of the component wording and the qualitative statements at the end of each page. The suggestions for changes in wording, movement of components between dimensions and merging of dimensions produced the third round survey. The resultant dimensions were called: complex physical terrain; complex human terrain; operational partnerships; socio-political issues; the nature, diversity and intensity of the threat; and the limits, capabilities and options of forces.

For each of these six dimensions, there were statements in the qualitative responses that needed to be assessed by the entire participant group before they could be considered to be supported. This was both because those who had not made the statements may not agree with them and because those who had made them may not agree with the researchers rewording to capture what was thought to be multiple repetitions of the same theme using different words. Within complex physical terrain, fifteen such statements were identified. For complex human terrain thirteen were identified. There were seven for operational partnerships, six for socio-political issues, sixteen for the nature, diversity and intensity of the threat and seven for the limits, capabilities and options of forces.

Nine such statements had been identified in the first round and presented in the second so that the level of general agreement could be ascertained. The results are shown in Table 6.

3.3.7. Design of the third round

The responses to the second round allowed construction of a nine page online third round survey. The first page asked for the rank of the participant. The next six pages each dealt with one of the six dimensions. Detailed descriptions of all proposed components were presented for each dimension. After the description, as in round two, participants were asked for a “yes”, “no” or “unsure” response to the two statements “do you agree that the concept expressed in this component should be included in this dimension” and “do you agree with the way this component is worded” as well as for their comments and suggestions. Interspersed with these questions were statements that came from comments made on the components in the second round. Participants were asked to rate their agreement with these statements on a 5-point Likert scale (“Strongly Disagree” “Disagree” “Neither Agree nor Disagree” “Agree” “Strongly Agree”). Following the questions on components, participants were asked for a “yes”, “no” or “unsure” response to the two statements “are there components that should be in this dimension but are missing” and “are there components that should be merged or split”. Those who answered “yes” were asked to describe the changes that they required. Finally, there was a question presenting comments relating to that dimension that has been made in the previous round and asking participants to rate their agreement on the same 5-point Likert scale used previously. Following the pages relating to specific dimensions, participants were asked to rank the dimensions in their order of importance for developing tactical, operational and strategic level scenarios. At the end, participants were asked for a “yes”, “no” or “unsure” response to the questions “are there dimensions that should be merged or split” and “are there important dimensions that are missing” as well as given an opportunity to make final comments.
tactical level the exclude for a simpler, reduced dimensional scenario. These are important for anyone building scenarios based on this dimensional framework as they show which dimensions to consider as it was not expected that they would. In place of such an overall question, participants were asked, for each component, whether any important dimensions are missing from this set, 21 experts responded yes and 0 were unsure; a consensus achieving (95.8%>87%) level of concurrence. Questioned on whether any important dimensions are missing from this set, 21 experts responded yes, 0 yes and 3 were unsure; again a consensus achieving (87.5%>87%) level of concurrence.

Participants also ranked the importance of these dimensions, and their underlying components and elements, at the strategic, operational and tactical levels. These ranks are given in Table 7 as is the mean rank given by participants. The mean rank can be seen as a measure of the distance between dimensions. As the rank does not tell the entire story the mean rank given by all participants is also shown. These are important for anyone building scenarios based on this dimensional framework as they show which dimensions to exclude for a simpler, reduced dimensional scenario.

At all three levels “the Nature, Diversity and Intensity of the Threat” was ranked as the most important dimension. At the tactical level the “Limits, Capabilities and Options of Forces” achieved a mean rank of 6.04 because one participant thought it so unimportant that it was ranked seventh out of six. Qualitative comments indicated that tactical scenarios are typically used to test the limits, capabilities and options of the force; and that this is, therefore, descriptive of the options (sometimes called strategies in scenario analysis literature) that are being tested by scenarios rather than a dimension of these scenarios. In the opinion of the respondents, the “Limits, Capabilities and Options of Forces” makes sense as a dimension at the operational and strategic levels. At these levels legacy systems, structures and procedures will always need to be dealt with. Interestingly, the “Complex Physical Environment” is seen as the least important dimension at the operational and strategic levels and only the third most important at the tactical level.

Unlike the dimensions, the issue of whether the entire group of components achieved consensus support as a complete set was not considered as it was not expected that they would. In place of such an overall question, participants were asked, for each component, whether they agreed that the concept (i.e. sub-theme) expressed in the component should be included and whether they agreed with the way the concept was worded. They were also given a chance to comment on how the wording of the component, or its title,

Table 6
Group agreement with controversial or important individual statements made in the first round. Note that SD is “strongly disagree”, D is “disagree”, NAD is “neither agree nor disagree”, A is “agree”, SA is “strongly agree” and NA is “no answer”.

<table>
<thead>
<tr>
<th>Statement</th>
<th>SD</th>
<th>D</th>
<th>NAD</th>
<th>A</th>
<th>SA</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>The climate and weather considered should be biased to our region.</td>
<td>4</td>
<td>9</td>
<td>6</td>
<td>11</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>The climate and weather considered should be representative of all possibilities.</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>19</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>It is important to consider complex urban terrain.</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>Appropriate interactions with other government agencies related to deployment (e.g. PM&amp;C) and return (e.g. AQIS) should be included in all scenarios.</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>We should always prepare to fight against adaptive, well led, competent, well trained, highly motivated, well supplied and (relatively) well equipped (for the nature of the conflict) enemies.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Decompression periods and locations need to be worked into strategies developed to cope with scenarios.</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>12</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>In all scenarios we should face an adaptive enemy who will alter the type of warfare to maximise their chances of meeting their objectives.</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>Lethality will continue to increase.</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>11</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>The ADF needs to ensure that a sufficient percentage of soldiers is reserved from deployment at any particular time that sufficient training can be undertaken to sustain the force.</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>19</td>
<td>13</td>
</tr>
</tbody>
</table>

3.3.8. Conduct of the third round
The third round commenced eleven days after the second round closed and was conducted as per the second round except that reminders were sent to all participants five days before the originally planned end which contained an extension of a fortnight. This extension was needed because of an overlap with a major Australian–US military exercise (Talisman Sabre 2009) in which many of the respondents participated. There were no formal dropouts. No students responded. The rank details of the 28 expert respondents are recorded in Table 2. Excluding those who left the survey window open for at least one entire night the time spent with the survey window open ranged from 14 min to 461 min with a median of 49 min.

3.3.9. Analysis of the third round
The consensus dimensions for the contexts in which the Australian Army may be deployed were: complex human terrain; complex physical environment; operational partnerships; socio-political issues; the nature, diversity and intensity of the threat; and the force limits, capabilities and options. Asked whether any of these dimensions should be merged or split, 23 experts responded ‘no’, 1 ‘yes’ and 0 were unsure; a consensus achieving (95.8%>87%) level of concurrence. Questioned on whether any important dimensions are missing from this set, 21 experts responded ‘no’, 0 ‘yes’ and 3 were unsure; again a consensus achieving (87.5%>87%) level of concurrence.

At the tactical level the “Limits, Capabilities and Options of Forces” achieved a mean rank of 6.04 because one participant thought it so unimportant that it was ranked seventh out of six. Qualitative comments indicated that tactical scenarios are typically used to test the limits, capabilities and options of the force and that this is, therefore, descriptive of the options (sometimes called strategies in scenario analysis literature) that are being tested by scenarios rather than a dimension of these scenarios. In the opinion of the respondents, the “Limits, Capabilities and Options of Forces” makes sense as a dimension at the operational and strategic levels. At these levels legacy systems, structures and procedures will always need to be dealt with. Interestingly, the “Complex Physical Environment” is seen as the least important dimension at the operational and strategic levels and only the third most important at the tactical level.

Unlike the dimensions, the issue of whether the entire group of components achieved consensus support as a complete set was not considered as it was not expected that they would. In place of such an overall question, participants were asked, for each component, whether they agreed that the concept (i.e. sub-theme) expressed in the component should be included and whether they agreed with the way the concept was worded. They were also given a chance to comment on how the wording of the component, or its title,

Table 7
Ranking of dimensions for the tactical, operational and strategic levels.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Tactical</th>
<th>Operational</th>
<th>Strategic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex physical environment</td>
<td>3rd (2.92)</td>
<td>6th (4.21)</td>
<td>6th (5.04)</td>
</tr>
<tr>
<td>Complex human terrain</td>
<td>2nd (2.46)</td>
<td>4th (3.67)</td>
<td>5th (4.38)</td>
</tr>
<tr>
<td>Operational partnerships</td>
<td>4th (4.38)</td>
<td>2nd (3.29)</td>
<td>3rd (2.88)</td>
</tr>
<tr>
<td>Socio-political issues</td>
<td>5th (4.63)</td>
<td>3rd (3.46)</td>
<td>2nd (2.21)</td>
</tr>
<tr>
<td>Nature, diversity and intensity of threat</td>
<td>1st (1.88)</td>
<td>1st (1.92)</td>
<td>1st (1.92)</td>
</tr>
<tr>
<td>Limits, capabilities and options of forces</td>
<td>6th (6.04)</td>
<td>4th (3.67)</td>
<td>4th (3.17)</td>
</tr>
</tbody>
</table>
Table 8
Support for retaining the concept expressed in each component in the dimensional hierarchy is expressed in ('yes'/‘no’/‘unsure’) form and those components reaching consensus are marked in boldface. Components with consensus support for their name and phrasing are marked in italics. Votes for whether each component is worthy of inclusion are given in Table 8 and those components that met the consensus criterion (>87%) on their name and phrasing are shown in bold in Table 8. Oddly, consensus was (>87%) for content are italicised. It is too confusing to also include the voting figures on the phrasing of components so those votes for whether each component is worthy of inclusion are given in Table 8 and those components that met the consensus criterion (>87%) are shown in bold.

<table>
<thead>
<tr>
<th>Complex physical environment</th>
<th>Physical Environment (27/0/0); Urban Development (27/0/0); Climate (26/1/0); and Infrastructure (25/2/0).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex Human Terrain</td>
<td>Population Characteristics (25/0/0); Population Allegiances (25/0/0); Heterogeneity (23/0/0); Complex Informational and Media Terrain (24/0/0); Language Skills and Barriers (24/1/0); Civilian Security (25/0/0); and Indigenous Capacity (23/1/0).</td>
</tr>
<tr>
<td>Operational Partnerships</td>
<td>Other Agencies (24/0/0); Interoperability (24/0/0); ADF Role in Partnerships (23/0/1); Contractors (23/0/1); Loss of Allied Presence (22/1/1); and Media Partnerships (22/2/0).</td>
</tr>
<tr>
<td>Socio-Political Issues</td>
<td>Host Nation Governance (22/0/2); Enemy Politics (22/1/1); Economic System (22/0/2); Support and Legitimacy (21/0/3); Endstates (23/0/1); Legal Issues (23/0/1); and Cycles (18/3/3).</td>
</tr>
<tr>
<td>Nature, Diversity &amp; Intensity of Threat</td>
<td>Nature, Diversity &amp; Intensity of Threat: Nature and Characteristics of Conflict (22/2/0); Nature of Enemy (24/0/0); Funding Resources (21/2/1); Threat Diversity (22/1/0); and Intensity (22/1/1).</td>
</tr>
<tr>
<td>Limits, Capabilities and Options of Forces</td>
<td>Limits, Capabilities and Options of Forces: Required Endstates (23/0/1); Force Design (23/1/0); Availability of Resources and Assets (23/1/0); Force Options (23/1/0); Duration (22/1/0); and Host Nation Capacity (17/4/3).</td>
</tr>
</tbody>
</table>

should be altered and whether it should be moved to another dimension or merged with another component. The ‘yes’/‘no’/‘unsure’ votes for whether each component is worthy of inclusion are given in Table 8 and those components that met the consensus criterion (>87%) for content are italicised. It is too confusing to also include the voting figures on the phrasing of components so those components with a consensus agreement (>87%) on their name and phrasing are shown in bold in Table 8. Oddly, consensus was (barely) achieved on the wording of the “Loss of Allied Presence” component even though there was a well argued case presented in the comments that this component should be called “Changing priorities of allies and partners”. Qualitative comments also indicated that several components should be moved out of their present dimension and merged with similar components under other dimensions. Strategic scenarios leave out the detail needed to construct operational scenarios or tactical vignettes for a good reason: there is a lot of it. Space prevents further exploration of the details of the comments on the names and descriptions of the components, let alone the elements.

Many factors impact on military deployments. In our case study, expert participants derived a consensus view about the most important of those factors. While the number of important factors remains large, they have been grouped into coherent and recognisable dimensions in a manner different to [1] but with similar outcomes. Within dimensions there are secondary coherent and recognisable groupings called components with these sub-themes themselves being composed of elements. The dimensions are not all equal in importance, as judged by panel members, and their relative importance varies with context (see Table 7).

3.4. Defining extreme dimensional outcomes

Most scenario methods define the extreme outcomes on their dimensions so that these can be used to frame the problem-space. This is the fifth step in Schoemaker’s ten-step process [4] and the fourth in Wright and Cairn’s eight-step process [1]. The extreme outcomes described in this section are based on a thematic analysis of the results by two of the authors and have not been represented to the participants for their comment. As such they need to be treated as hypotheses about what the extreme outcomes are.

Physical environments have always been important in warfare, with factors such as the slope of the ground, the position of a river, or the use of a mountain pass being critical to the outcomes of historical battles. The first dimension, Complex Physical Environment, extends the natural environment by including human changes, from trenches and tunnels to urban development and infrastructure. Theming based on the qualitative comments on the four components discussed in Table 8 and responses to sixteen specific questions presented in the third round based on second round qualitative comments lead the authors to hypothesise that the extreme values of this dimension would be “cluttered” and “clear”, as shown in Table 9. These could equally be expressed as ‘close’ and ‘open’ terrain. Surviving on the lethal modern battlefield requires use of cover and concealment. Cluttered terrain is problematic because your enemies have cover and concealment and clear terrain is problematic because you do not. Some respondents made forceful arguments for the inclusion of urban terrain as a special case but it can be seen as cluttered terrain with integrated civilians.

Just as aspects of the physical terrain can protect a force or make it vulnerable, so can aspects of the human terrain of the host nation. Complex Human Terrain covers characteristics and allegiances of the population, their heterogeneity and their languages, as well as the informational and media terrain that they create. It also covers the level of civilian security and the effectiveness of other government services. Theming qualitative comments from the seven components discussed in Table 8 and thirteen specific questions led the authors to the hypothesis that “integrated” and “separate”, see Table 9, are the extreme values of this dimension. The responses made it obvious that warfare is conducted differently in a space shared with civilians than in isolated regions. However, integrated does not simply mean fighting with civilians on the battlefield but fighting in such a place where your enemy is integrated.

Table 9
Extreme outcomes for dimensions.

<table>
<thead>
<tr>
<th>Complex physical environment</th>
<th>Cluttered</th>
<th>Clear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex human terrain</td>
<td>Integrated</td>
<td>Separate</td>
</tr>
<tr>
<td>Operational partnerships</td>
<td>Junior</td>
<td>Senior</td>
</tr>
<tr>
<td>Socio-political issues</td>
<td>Favourable</td>
<td>Unfavourable</td>
</tr>
<tr>
<td>Nature, diversity &amp; intensity of threat</td>
<td>Modern-system state</td>
<td>State supported insurgency</td>
</tr>
<tr>
<td>Limits, capabilities and options of forces</td>
<td>Regular only</td>
<td>Regular plus</td>
</tr>
</tbody>
</table>
into at least a part of that civilian population and can conceal itself amongst them. Separate combat may still occur close to civilians, but separation of both forces from civilians means discrimination thresholds will be crossed at a greater range.

The opposing forces in a war do not fight in isolation from the rest of the world. This is covered by the Socio-Political dimension, which has a wider coverage than its name implies and deals with factors that are often classified as strategic. How healthy are the economic systems of the threat and host nations? What level of support and legitimacy does the conflict have? Is there likely to be an impact from global economic cycles? What legal issues are raised by the conflict? Theming qualitative comments from the seven components discussed in Table 8 and seven specific questions led to the hypothesis that socio-political factors that are “favourable” or “unfavourable” are the extreme values of this dimension, see Table 9.

Wars also require the effective management of many partnerships. A national army is likely to have allies or operational partners and will need to deal with contractors and government agencies from several countries and with international organisations. This is covered in the dimension Operational Partnerships. The results emphasise the degree to which the officers consulted valued the increased mission effectiveness that can be achieved by better integration with partners, be they joint (Navy or Air Force), other government agencies (e.g. Australian Federal Police or AusAID), host nation agencies, allied forces or non-governmental organisations (e.g. Red Cross or MSF). The presence of this dimension supports the increasing trend towards involving representatives of these organisations in exercises that use scenarios. Qualitative thematic analysis of the comments on the six components in Table 8 and seven specific questions based on qualitative responses in the previous round led to the hypothesis that the ADF would be a “junior” or a “senior” partner (see Table 9). As a junior partner it is likely that many aspects of sustainment, health, command, control, communication, computer, intelligence, surveillance and reconnaissance will be undertaken by the senior partner who the ADF will have to integrate with. As a senior partner the ADF would need to undertake these tasks for junior partners.

Participants considered the Nature, Diversity and Intensity of Threat as the most important dimension at all levels (see Table 7). This covers the nature of the potential conflict, the capabilities, size and organisation of the threat force, the source of funds for the threat force, the possibilities of more diverse threats, e.g. across borders, and the likely intensity of conflict. Theming of qualitative comments on the five components (see Table 8) and responses to sixteen specific questions based on second round responses lead the authors to hypothesise that the extreme values of this dimension would be facing a “modern-system state” adversary [72,73] and a “state-supported insurgency”, as shown in Table 9. These appear to have been the two extremes because a well handled conventional force requires force concentration on a theatre wide basis whereas counter insurgency operations require local concentration of combined arms firepower but a more dispersed force across the entire theatre. The respondents indicated that opponents should be Chemical, Biological, Radiological and Nuclear (CBRN) capable to some extent, because these capabilities make fighting them harder. They also indicated that emergency relief, peacekeeping, defence aid to the civilian community, services assisted evacuation, services protected evacuation, sanctions enforcement, peace enforcement, counter terrorism and counter WMD proliferation should be woven into the scenarios.

When gathering a force to combat a threat, it might be desirable to select deployed units based on criteria such as readiness, capability against the threat and compatibility with other chosen units; but some desirable units may already be committed to other campaigns, certain specialists may be in short supply, particular units may be held in reserve for the security of the nation and there will be limits on the budget. These factors are covered in the dimension Limits, Capabilities and Options of Forces. Qualitative analysis of the comments on the six components in Table 8 and seven specific questions based on qualitative responses in the previous round led to the hypothesis that the extremes of this dimension are “regular only” and “regular plus” (see Table 9). In a small war or for a small number of deployments the regular army would be able to provide the troops needed in a deployment as part of its normal tempo of activities. Coping with this option stresses the Army systems by forcing them to balance readiness and training within the regular Army to meet requirements for present and future deployments. In a larger war, or for numerous simultaneous deployments, the Australian Army would need to call upon reserves and possibly even recruit soldiers rapidly from civilian life or contract out certain specialist tasks (e.g. construction design work). An option that expands deployable forces beyond the regular Army stresses Army systems by making them consider how to integrate reservists and contractors into deployments.

3.5. Applicability of the dimensions

The original expert panel found these six dimensions to form a plausible set for the contexts in which the Australian Army may be deployed.

4. Discussion

Strong willed and knowledgeable participants were important to the success of this process. The majority of participants were senior enough to be eligible to be commanders of battlegroups deployed overseas, with staff officers to advise them and responsible for the lives of many hundreds of soldiers. Their training and personality predispose them to being critical and obsessed with details while keeping the larger picture firmly in mind. They were extremely critical of the candidate dimensions that we presented to them at the start of the first round. In line with [54] we selected 87% ‘Agree’ or ‘Strongly Agree’ as a consensus level, but only 9% (3 of 34) candidate dimensions, ‘human geography’, ‘shaping the battlespace’ and ‘enemy information operations’, met this criteria. This low rate of consensus and the qualitative comments led to a fundamental revision of the dimensional structure. Themetic analysis was used on the qualitative responses to construct a nested structure of seven dimensions each containing a number of components. One of the strengths of Delphi is that researchers submit their interpretations of the qualitative responses of participants, and the thematic structure drawn from them, back to the judgement of those participants. This was done in the second round and 68% of the
components met the consensus threshold as did six of the seven dimensions. The qualitative responses from this round were used to alter and shift components and remove one dimension. When shown to the participants in the third round, consensus was achieved on 94% (34 of 36) of the components and all of the dimensions. Of the two non-consensus components one, “commitment limitations”, with 70.8% support, was removed. Another, “cycles” also had 70.8% support yet was retained but clearly marked. This partial retention was due to the strong support given in some comments and the ability to modify the descriptive text in a manner consistent with both positive and negative comments. The three components that barely attained consensus, all with 87.5% support, did so due to a high level of ‘unsure’ responses rather than outright rejections. These ‘unsure’ responses were linked with comments suggesting minor alterations to wording. These minor changes were made and the three components, “support and legitimacy”, “funding” and “host nation capacity”, were included in Table 8.

4.1. Validity and reliability of results

Of the ninety-three senior and field grade officers in the Australian Army employed to think about future land warfare at the time of this study, or who had published on it in a peer-reviewed journal over the previous three years, seventy-three participated. Their responses were detailed and had considerable consistency amongst their diversity. The Delphi process we used assessed participant agreement with the constructs presented in each round quantitatively and used qualitative analysis of responses to alter the constructs prior to the following round. An 87% consensus threshold was used and achievement of consensus is seen as evidence of validity [54], as the experts have agreed on both the concepts and their wording. The higher the consensus threshold, the more evidence of validity. All dimensions and thirty-one of the thirty-five components listed in Table 8 met the 87% consensus threshold, and the four that did not are clearly marked.

Delphi studies on areas for which there are large numbers of experts, or in which little expertise is required to respond cogently, often use sophisticated statistical methods to indicate the applicability of their outcomes. Only twenty-eight of the initial seventy-three expert participants responded to the third round but this dropout rate is consistent with three round Delphi processes [74]. The participants individually displayed deep expertise and experience and, as a group, considerable breadth across the totality of military domains, the critical factors for this study. The quality of Delphi results is tied to the quality of the panel rather than its size [54,75].

4.2. Participant rankings of dimensions

The participant’s ranking of the importance of the dimensions in building tactical, operational and strategic level scenarios are shown in Table 7. The mean rank given by participants, shown parenthetically in Table 7, reveals that the distances between dimensions are non-constant. This ranking allows people building scenarios based on this dimensional framework to understand which dimensions the participant’s thought that they should exclude for a simpler, reduced dimensional scenario.

At all three levels “the Nature, Diversity and Intensity of the Threat” was ranked as the most important dimension. At the tactical level the “Limits, Capabilities and Options of Forces” was not seen to make sense and achieved a mean rank of 6.04 because one participant ranked it seventh out of six and the rest all ranked it sixth. “Complex Physical Environment” was seen as the least important dimension at the operational and strategic levels and, while just 0.46 of a rank away from second in mean rank terms, was only the third most important dimension at the tactical level.

The mean rank results can be seen as a tip-off as to how many dimensions are needed to describe the problem-space. A large gap in mean ranks implies that there is a consistent participant view that there is a large gap in importance between two dimensions. Thus it may make sense to use the top three dimensions to describe tactical scenarios because of the jump from a mean rank of 2.92 to 4.38 between the third and fourth ranked dimensions, yet the closeness of the second to sixth mean ranks at the operational level may indicate that scenario analysts should jump from a one dimensional scenario-space (as the ACCS, which are really operational level scenarios, actually use) to a six dimensional scenario-space.

5. Conclusion

We have presented a method integrating thematic analysis between Delphi rounds to define a dimensional hierarchy of important variables at the strategic, operational and tactical levels. This online technique facilitates expert participation by allowing asynchronous access from many separate locations and through the use of thematic analysis to remove time consuming duplicate argumentation and standardise language conventions. Thematic analysis also builds a hierarchy of strategic dimensions, operational components and tactical elements to be re-presented to experts in the subsequent Delphi round. Anonymity and thematic de-duplication allow merit-based discussion by masking the status and numbers of advocates. In our case study, experienced Australian Army officers committed to force development worked with us to develop an explicit statement of the dimensions of future land warfare for the Australian Army: complex physical environment; complex human terrain; operational partnerships; socio-political issues; the nature, diversity and intensity of the threat; and the force limits, capabilities and options. These dimensions provide a draft statement of Army requirements for future ADF scenario reviews. They have also been used to assess the suitability of sets of scenarios and to compare sets of scenarios to each other and to a benchmark set (the ACCS). Experts were able to contribute as much as they felt able to at a time and place of their choosing and were heard based on their arguments rather than their hierarchical position. The median participant spent slightly less than 3 h on the entire process. There was no travel time and no time spent waiting for those with more extensive contributions. Participants represented their positions in writing, thus eliminating errors from researcher transcription and saving enough time to allow analysis and construction of the next round to occur.
in a fortnight. Iteration through rounds of comment and quantitative voting on themes derived by researchers from the written participant responses in the previous round allowed experts to assess how accurately analysts themed their responses as well as how much they agreed with the responses from the entire group. Consensus was achieved after three rounds for the dimensions but the case study was not continued to achieve consensus for operational components or tactical elements. Both the dimensions and the technique are novel. The dimensions are specific to the Australian Army and would require joint inputs to ensure relevance to the entire ADF. However, the technique is generalisable to other organisations with an interest in robustly preparing for the future, particularly those with an entrenched dominance hierarchy and experts who are busy. Furthermore, it may be used as an online dimensionalisation stage before an online scenario development method.

Acknowledgements

Profs. Jan Pincombe and Drew Dawson advised on methodology. LTCOLs Andrew Fullgrabe and Pat Sowry, the DSTO authors and Mrs Margaret Egudo were members of the panel of experts who refined the questions before the validity stage. LTCOL Andrew Fullgrabe recruited participating unit members. Brigadier Mal Rerden, COL John Hutcheson and LTCOL Pat Sowry recruited the main study participants. Mrs Margaret Egudo helped theme the first round. This study was sponsored by Brigadiers Justin Kelly, Gus Gilmore and Mal Rerden and COL Damian Cantwell as Director General of Future Land Warfare (DGFLW). Access to participants was provided by Brigadiers Mal Rerden, Shane Caughey, Simone Wilkie and Gus Gilmore. LTCOLs Damian Cantwell, Peter Steel, David Wainwright, Dave Buller, Trent Stothart, Chris Mills and Pat Sowry served as desk officers and Dr Darryn Reid, Mr John Coleby and Mr Dean Bowley as task leaders. Aubyn Pincombe critiqued the manuscript. Dr Darryn Reid, Dr Neil Reid and BRIG (Ret.) Steve Quinn reviewed this paper within DSTO in their roles as head of (land) long range research, leader of (land) operations research and chief of Land Operations Division respectively. We gratefully acknowledge the contributions and efforts of all participants. Ranks and positions are as they were when assistance was given.

References
