The Science Inside

How Can Dstl Expand Our National Security Gaming Toolset To Generate More Meaningful And Reliable Insights?

Bennett, J, Bagwell, M, Ho, S, Strong, P

DSTL/PUB131779 1.4
28 May 2021

Dstl
Exploration Division
Portsdown West
Portsdown Hill Road
Fareham
PO17 6AD

Release Conditions

Authorisation		
Role	**Name**	**Date**
Project Manager	Lesley Pyne	02/12/2020
Technical Reviewer	Andy May	02/12/2020
Author	James Bennett	02/12/2020

Executive summary

Introduction

The national security games Dstl have conducted in recent years have added value in understanding Blue processes, tools and capabilities, but have fallen short in areas such as engendering truly strategic dynamics, understanding possible Red responses to Blue actions, and how responses to crises might play out. Furthermore, to fully address the questions that are being asked of us by senior Her Majesty's Government (HMG) stakeholders we have identified a requirement to generate more meaningful and valid insights from our national security games. The authors assert that increasing the level of challenge and analysis in these games will serve to generate more meaningful and valid insights, and this report outlines some potential solutions as to how this can be accomplished.

How Is An 'Analytical Game' Defined?

We define an analytical game as:

A game that employs analytical approaches and/or methods to generate insights as part of an analytical process.

The goal of an analytical game is to *create* rather than to *convey* knowledge. We consider games that create knowledge to be those which are undertaken as part of an analytical process, whose purpose is gaining insights into a problem space.

How Can We Develop Creating Knowledge Games That Are More Analytical?

We have identified a number of ways in which we can approach the design and development of our games more analytically. These include:

- Engaging in more thorough requirements capture and working with sponsors to set research questions that are limited in number and scope, and more appropriate to the sponsor's objectives.
- Making data capture more central to game design and fully documenting design choices.
- Ensuring that game design choices are driven by effective rigour and the game's analytical requirements, rather than being dominated by considerations of player engagement and immersion.
- Conducting effective validation and verification using the Evidence Framework Approach.
- Better employing analytical methods during post-game analysis.

How Can We Conduct More Analytical Games Within The Constraints Of Engaging Very Senior Players?

Conducting gaming with senior players can offer invaluable insights into how senior decision-makers might perceive and respond to issues. However, engaging senior players imposes a number of constraints on game design, affecting, amongst other things, the way in which requirements are set and how sponsors oversee the design, the time available for execution, the game mechanics that can be used, and the extent of the game design team's ability to direct player actions. Given these constraints we recommend only involving senior players where this is necessary for achieving primary game objectives. Where senior players are considered essential to answering primary research objectives, consideration should be given to how best to maximise the value of their inputs; they may not need to be present at every part of the process.

Analytical approaches to gaming require a rebalancing of player immersion and mechanics that generate game outputs suitable for analysis. This can mean that a 'bought in' sponsor or senior representative should be present to ensure compliance. We also need to provide a degree of education on gaming for sponsors and stakeholders who often lack experience in commissioning, attending, and making use of the results of games.

How Can We Encourage More Representative Red Cell Responses To Blue Cell Actions?

In order to improve the ability to generate more meaningful and valid insights it is important to ensure that the adversary is adequately represented, and we have identified a number of challenges in representing Red. The Red Cell has often been a comparatively neglected element of game design, as sponsors generally prioritise exploring Blue decision-making and the UK and Allied processes for formulating adequate responses to crises. Red Cells have also lacked independence from the Control Cell, with their actions being constrained in order to ensure that they produce actions most suited to allowing the Blue Cell to explore issues and generate insights in line with the game objectives. Useful representation of Red/Blue interactions can often by hindered by the provision of game starting conditions and mutually exclusive and simplistic player objectives which inevitably drive players towards conflict.

To combat this we recommend that Red Cell mechanics and objectives should encourage a more competitive spirit and desire to succeed. This competitive spirit should be tempered by realistic objectives and constraints. Their objectives should always be more nuanced than simply defeating Blue and should be grounded in as thorough an understanding of the real-world adversary as possible. They should also leave open the possibility for modification or reprioritisation to account for circumstances or opportunities arising within the game. Game designs should also rebalance their focus, placing more emphasis on creating Red cells that also involve more complex interactions between competing internal elements. A degree of Control Cell direction to ensure that Red Cell play meets game objectives will still be necessary; however, Red Cells must be given more freedom to respond to Blue actions as they see fit.

Conclusions and Recommendations

This paper makes an extensive number of recommendations regarding the characterisation of analytical national security games, the requirements capture and design process, game methodologies and design features, sponsor and player engagement, and data capture and analysis. All of these recommendations seek to improve the ability to conduct analytical games on national security issues within the MOD and across Her Majesty's Government.

We believe that this paper should be part of an ongoing process of assessment and evaluation of the methods we use in our national security games. We also recognise that there many gaps in this paper that would require further exploration in future work. We therefore recommend that this research be periodically revisited and updated to account for improvements in our own knowledge and experience, and to incorporate best practice from our colleagues in other governmental organisations in the UK and amongst our allies, industry and academia.

Table of contents

List of tables

List of figures

1 Introduction

1.1 Research Requirement 1

In 2017, Dstl were tasked by the then-Vice Chief of Defence Staff (VCDS) to make recommendations on how wargaming could be reinvigorated in Defence. One of the recommendations made was to inculcate a culture of wargaming through senior level sponsorship and active participation in wargames. One of the activities to flow from this recommendation was the VCDS series of senior strategic wargames.[1]

While Dstl has a lengthy track-record of delivering senior Table Top Exercises (TTXs), this series of games represented the first time in many years that dynamic, interactive games on national security issues had been conducted in the United Kingdom (UK) with serving senior personnel. The national security games we have conducted so far have added considerable value in identifying the challenges the UK faces and driving action to address them. Whilst that the games we have conducted so far have added value in understanding blue processes, tools and capabilities, we believe that they have fallen short in areas such as engendering truly strategic dynamics, understanding possible red responses to blue actions, and how responses to crises might play out.

There is a growing interest in the Ministry of Defence (MOD) and across Government in the use of gaming to explore strategic issues in complex wicked/unstructured problem spaces. There is a recognition amongst sponsors and stakeholders that gaming could be used to provide insights into strategic questions[2] where existing empirical and deductive methods have been found wanting, or have not been attempted. To fully address the questions that are being asked of us by senior Her Majesty's Government (HMG) stakeholders we have identified a requirement to generate more meaningful and valid insights from our national security games. Lessons identified from previous national security games have also highlighted where design choices to meet sponsor requirements and overcome the constraints of conducting games at the senior level – exemplified by a lack of time/availability, alongside scepticism of gaming methods – have previously precluded the development of more analytical techniques.

These observations come against a backdrop of numerous reviews which have advocated for greater challenge to decision-making and care regarding the

[1] In the past we at Dstl have internally referred to these types of games as either 'strategic wargaming' or 'strategic gaming'. However, the authors believe the term 'national security games', more accurately reflects the activities we undertake. Elizabeth Bartels, whose work we draw heavily on in this paper, champions the use of the term 'national security gaming' rather than 'wargaming' or strategic gaming because it '…denotes a broader range of topics that better reflect the actual application of the tool to issues such as crisis management, measures short of war, and bureaucratic policy areas such as acquisitions and personnel that have major implications for national security beyond fighting major wars. It also is more inclusive of diplomacy and development community members who may find the term "war" off-putting culturally as well as not being descriptive of their work'. Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, 2020, Pardee RAND Graduate School, p.2

[2] Such as deterrence, escalation, assurance, strategic communications, de-escalation, off-ramps, and re-establishing deterrence and so on.

understanding and use of evidence. Lord Levine's 2011 UK Defence Reform report argued for the need for greater evidence-based decision-making while noting the danger that evidence can be misused to support advocacy rather than objective decision-making[3]. The 2015 Review of Analysis by Roger Hutton argued for further reforms to increase the centrality and proper use of analysis in decision-making[4]. The 2016 Iraq Inquiry raised issues concerning the UK's evaluation, assessment and understanding of evidence[5]. In particular it highlighted the value of multiple perspectives and dissenting views, the need "to be scrupulous in discriminating between facts and knowledge on the one hand and opinion, judgement or belief on the other", and the "need for vigilance to avoid unwittingly crossing the line from supposition to certainty, including by constant repetition of received wisdom"[6]. Given the above the authors assert that there is a need to increase the level of challenge and analysis in Dstl's national security games, and that doing so will serve to generate more meaningful and valid insights. The intent of this report will therefore be to provide some potential solutions as to how this can be accomplished within the constraints under which our national security games are required to operate.

The primary requirement for this research has been driven by a need to address the lessons identified from previous national security games and fill the gap in our analytical gaming methods.

Key lessons to be addressed include:

a. **Experiential Value:** The primary value of senior level national security games has thus far been from the participants' experiences of decision making and living with difficult choices taken during gameplay – they identified / discovered new issues, derived fresh insights about the challenges of implementing responses, and tested ideas and perspectives amongst their peers. Whilst useful, it is hard to ensure that participants' self-generated insights and lessons are a reasonable reflection of the game design, and take proper account of the assumptions that underpin the game mechanics, scenario and data used in the game.

b. **Identifying Genuine Insights:** We argue that genuine insights from analytical games are those which are the product of appropriate analytical methods that have been applied to the data which was captured to draw conclusions that are both valid and have been validated against an objective standard. They are also those which relate to the aspects of the problem-space that the game was designed to examine. By contrast, in some previous cases the 'insights' that players drew from the game were clearly a direct product of the scenario/mechanisms baked into the game design by the designers; this was mainly a cause for concern in relation to insights drawn regarding the understanding of Red actions, behaviours and responses towards Blue. It has long been an issue that players and sponsors find it difficult to separate aspects of the game design and scenario, which are used to elicit game outputs, from the intended outputs themselves. This can lead to players and sponsors incorrectly

[3] Levene, 2011, "*Defence Reform: An independent report into the structure and management of the Ministry of Defence"*
[4] Hutton, 2015, *Review of Analysis in Defence: Interim Report*
[5] HC264, 2016 *The Report of the Iraq Inquiry: Executive Summary*
[6] *Ibid,* Executive Summary, pp.131-132

drawing conclusions relating to game inputs rather than outputs, and misinterpreting and overstating the meaning of game outputs. This was encapsulated in a famous exchange between Robert Levine, Thomas Schelling and William Jones, all RAND Corporation analysts in 1964, in which Levine stated:

> *"[Games'] excitement and the logical problems of structure they present seduce those who intend to use them economically into using them elaborately and frequently; their surface plausibility seduces those who enter them sceptically, "merely looking for hypotheses," into leaving them with conclusions.[7]"*

c. **Post-Game Analysis:** Comparatively few genuine insights have been derived from post-game analysis of data captured during gameplay, which can often provide deeper, and more robust insights than the initial impressions generated by the players themselves. The recognition of this issue leads to a requirement to directly address what sort of genuinely useful analysis can be undertaken on highly qualitative discussion based games, and what value would this add over our current more straightforward approach based primarily around writing up the narrative.

1.2 Research Requirement 2

As deterrence, coercion and escalation represent a major subset of the areas of focus for our national security games, we considered that developing a proof-of-principal game against one of the candidate questions would be a useful complement to Requirement 1. Developing such a game would allow us to explore the practical utility of the recommendations made in this paper and would provide us with a tangible demonstration of what a game designed on analytical principles could look like and the benefits of adopting such an approach.

1.3 Research Questions

In line with the above the following overarching questions/problems have been identified that this report will need to address:

a. How is an analytical game defined?
b. How can we develop 'creating knowledge' games that are more analytical?

Research question c addresses the particular practical problems we have encountered when running national security games with very senior stakeholders and players:

c. How can we conduct more analytical national security games within the constraints of engaging very senior players?

[7] Levine, Schelling, Jones, "Crisis Games 27 Years Later: Plus C'est Déjà Vu", p.1

Research question d is based on our assessment that the gameplay dynamics between cells proved to be either unrealistic/unrepresentative or not amenable to deriving post-game insights:

d. How can we encourage more representative Red cell responses to Blue cell actions?

1.4 Proof of Concept Games

This report has been written in parallel to the design and development an analytical proof of concept game. This game broadly addresses elements of the questions outlined above, and so drew from this research as it progressed, as well as providing further contributions to the report's conclusions. The games was designed to address the questions in as optimal an environment as possible, i.e. one which lacks the constraints that have previously led to significant compromises in Dstl's game designs, which include; problematic sponsor relationships, poorly-defined requirements, competing objectives, and compressed development timeframes. These constraints are covered in more detail in section 4.

The intent of this game was to allow us to understand what a more analytical game might look like and what it would entail from a design perspective – it is a representation of best practice that will help us to identify aspects which would could apply in future customer-facing work.

Alongside addressing the questions proof of concept games will provide Dstl's national security gaming capability with additional value, as they will allow us to:

- Provide demonstration games to show future stakeholders what additional value can be achieved when constraints are lifted or adequately mitigated;
- Design, implement and test methods to enhance the fidelity of data capture – thus creating the circumstances for higher quality post-game analysis to be undertaken;
- Allow us to draw conclusions relating to analytical games as to whether they are mechanically useful as well as being palatable to a senior audience and, if not, to propose mitigations/alternative methods.

1.5 Report Structure

This report contains the findings of our research into the question of how Dstl can expand our National Security Gaming toolset to generate more meaningful and valid insights. It answers the question above and presents a number of conclusions and recommendations as to how we should execute future games in this space.

This report is divided into a number of sections, each of which focuses on answering one of the research questions outlined above.

Section 2 – How Is An 'Analytical Game' Defined?

- Presents a series of taxonomical statements intended to define what an analytical wargame actually is as well as its key tenets, based on research and a literature review.

- Presents a typology for classifying the different types of analytical National Security game that we could be required to execute.

Section 3 – How Can We Develop Creating Knowledge Games That Are More Analytical?

- Provides an evaluation of how we undertake the gaming process. We present a more codified/structured approach to design and development that integrates validation and verification of the construct, as well as a number of improvements to our current methods of game design, data capture and post-game analysis.

- We put forward a schema for defining the questions and objectives of a game.

Section 4 – How Can We Conduct More Analytical Games Within The Constraints Of Engaging Very Senior Players?

- Outlines a range of game design constraints placed on us by sponsors and from engaging with very senior players.

- Provides a range of potential options to mitigate these constraints.

Section 5 – How Can We Encourage More Representative Red Cell Responses To Blue Cell Actions?

- Outlines the current challenges of representing Red in national security games.

- Provides a detailed assessment of the role(s) that Red can play in games.

- Provides a range of ways we could make our Red cells more representative, ranging from different cell composition to effective pre-game training.

- Assesses how we represent Red during game design and provides a number of example game designs that help to mitigate the issues identified throughout.

Section 6 – Proof of Concept Escalation Dynamics Game and Concept of Analysis

- Provides the outline design for a bespoke proof of concept game examining escalation dynamics.

- This game addresses many of the issues raised throughout the paper and provides an exemplar game design proposal.

Section 7 – Conclusions and Recommendations

- Provides a range of conclusions from the research undertaken during this report.

- Provides a range of recommended actions we should take during future national security games which should help to address the range of issues highlighted throughout the paper.

Section 8 – Closing Summary

2 How Is An 'Analytical Game' Defined?

2.1 Terminology: 'Game' Defined

Throughout extant gaming literature the term 'wargame' has been given a variety of definitions by different practitioners.[8] Throughout this document we use the term 'game' in generally the same manner as most practitioners would use the term 'wargame', but we have made a purposeful distinction between the two to delineate that our games do not focus specifically on war as their main area of interest.

When referring to a 'game' this work will use a modified version of the definition of a 'wargame' put forward by Graham Longley Brown, co-author of the Ministry of Defence's *Wargaming Handbook*, in his most recent publication *Successful Professional Wargames*:

> *Adversarial and oppositional by nature, a game is an immersive simulation of real-world activities within an environment, in which the course of events affects, and is affected by, decisions made by the players*.[9]

2.2 Conceptual Archetypes of Professional Games

Conceptually, the senior national security games Dstl have run up until this point can be categorised as 'Discovery Games', intending to create knowledge – i.e. their goal was to generate novel insights into a strategic problem. Table 1, below, shows a number of archetype created by Bartels which shows discovery games and 'Analytical Games' as those which create knowledge:

[8] Perla, *The Art of Wargaming,* 1990, p.164, Sabin, *Simulating War: Studying Conflict Through Simulation Games*, p.4, DCDC, 'Wargaming Handbook', 2017, p.5.
[9] Longley-Brown, *Successful Professional Wargames: A Practitioner's Handbook*, p.46.

The modification was to change 'not involving the operations of actual forces' into 'not involving real world activities'. This is to reflect the fact that many wargames Dstl runs are not focused on war or warfighting, which is somewhat implied in the original texts used of the word 'forces'. This change broadens the definition to clearly include activities outside of war/warfighting.

		Goal of Game	
		Creating Knowledge	**Conveying Knowledge**
Problem Type	**Unstructured/ Wicked Problem[10]**	Discovery Games	Educational Games
	Structured Problem	Analytical Games	Training Games

Table 1 - The four overarching archetypes that comprise professional games

Bartels states that Discovery Games 'seek to generate new hypotheses and variables of interest about an unstructured problem. Games of this type can include investigations of new problems, consideration of potential future conditions and rare events such as black swans'.[11] We would also add that these games are useful for exploring poorly understood problems in unstructured spaces. The execution of games of this type has provided substantive value to both stakeholders and the game's participants, due to their experiential nature and the lessons learnt by participants during gameplay.

In introducing the archetypes above, Bartels defined Analytical Games as those which 'seek to help us understand phenomena that are relatively well structured. Thus, these games often seek to illustrate how human decisions interact with well-defined (often physical) phenomena, like weapons system performance'. Since most of the strategic problems we face are not so well-defined and structured, national security games would tend not to fit into this category. However, we, and indeed Bartels in her later work, contend that this definition is too narrow, and it is indeed possible to take more analytical approaches to national security problems.

It is important to note that there can be an overlap between Creating Knowledge and Conveying Knowledge games – games designed for one purpose might also have benefits in the other[12]. For example, a game designed primarily for creating knowledge by examining a problem might also impart some knowledge regarding the issues concerned to the participants. However, as Francis McHugh wrote in 1966, and has been quoted in the UK's Defence Wargaming Handbook, "it is better to point the game towards but one of these objectives[13]". In other words, games should be designed primarily for only one of these purposes rather than explicitly trying to

[10] An unstructured/wicked problem is defined in Rittel and Webber in "*Dilemmas in a General Theory of Planning*." Policy sciences, 4(2), 1973, pp.160-167. Wicked problems have several important characteristics: 1) They do not have a definitive formulation. 2) They lack an inherent logic that signals when they are solved. 3) Their solutions are not true or false, only good or bad. 4) There is no way to test the solution. 5) They cannot be studied through trial and error. 6) All wicked problems are essentially unique. 7) The way a wicked problem is described determines its possible solutions.

[11] Bartels, "*Innovative Education: Gaming - Learning at play*", 4 Aug 2014, https://pubsonline.informs.org/do/10.1287/orms.2014.04.13/full/

[12] DCDC, 'Wargaming Handbook', 2017, Ministry of Defence, p.9

[13] McHugh, F., Fundamentals of War Gaming, US Naval War College, 3rd Edition, 1966, page 9, quoted in DCDC, 'Wargaming Handbook', 2017, Ministry of Defence

achieve both in the same game. The purpose of this report is to focus on games designed primarily for the purpose of creating knowledge.

2.3 Terminology: 'Analysis', 'Analytical' And 'Analytical Game' Defined

'Analysis' is defined as:

> '*a detailed examination of anything complex in order to understand its nature or to determine its essential features: a thorough study'.*[14]

'Analytical' is defined as:

> *'using a method of separating things into their parts in order to examine and understand them'…'an analytical way of doing something involves the use of logical reasoning'.*[15]

The MoD defines 'Operational Analysis' as:

> *'the application of scientific methods to assist executive decision makers'.*[16]

Although most national security games will continue to be Discovery Games, as the problem spaces they grapple with are almost exclusively unstructured and difficult to quantify, it can be argued that Discovery Games and Analytical Games[17] sit on opposite sides of a single analytical spectrum. Both have the potential to be analytical, but there is a natural tendency to increase the depth of analytical rigour when moving from the discovery space towards 'pure' Analytical Games. Such games that are designed to deal with bounded well-structured problems that are generally part of a more ordered domain within which we can potentially make more concrete conjectures relating to cause and effect.

Table 1 showed Bartels's conceptual archetypes, within which she defines an 'Analytical Game' as one which creates knowledge in the context of a structured problem space. We generally agree with Bartels's schema. However, since we argue

---

[14] Merriam-Webster Dictionary, 'Analysis', https://www.merriam-webster.com/dictionary/analysis

[15] Macmillan Dictionary, 'Analytical' https://www.macmillandictionary.com/dictionary/british/analytical
Collins Dictionary, 'Analytical', https://www.collinsdictionary.com/dictionary/english/analytical

[16] 'CSA's Review of Operational Analysis in Contributing to Policy, Planning and Acquisition Decision Making', July 2004.
This definition more explicitly links analysis to the purpose for which it is undertaken – i.e. to assist decision makers. This helps national security game designers to remember that they should always design to a purpose.

[17] Bartels states that 'I use "analysis" in the general sense to refer to efforts to better understand the elements and structure of a policy area… I use the term to apply to a broad range of activities that are also sometimes referred to as research, inference, or studies—such work need not be quantitative (as is sometimes inferred) nor devoted only to approaches that decompose aspects of a larger problem (as is indicated in some formal definitions of the term "analysis")
Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, 2020, p.3.

that all "creating knowledge" games sit on an analytical spectrum 'Analytical Games' as she defines them is not a helpful term, as it implies that other types of creating knowledge games – i.e. Discovery Games – are not analytical.

We would therefore argue that the definition of an analytical game is broader than that proposed by Bartels. For the purposes of this paper we will define an analytical game in the broadest possible terms as:

> *A game that employs analytical approaches and/or methods to generate insights as part of an analytical process.*

In this context 'analytical approaches and methods' encompasses a rigorous approach to the generation of inputs and derivation of appropriate game mechanics during game design to engender greater confidence in insights generated, as well as the potential employment of an entire range of qualitative and quantitative methods when analysing a game in order to draw some form of insight. Given this definition we would contend Discovery Games should still be considered analytical games, since their inputs allow them to discover genuine insights that were previously unknown to the game design team or sponsor, and qualitative methods of analysis can potentially be applied to the data they produce to generate insights.

We fully agree with and would like to stress Bartels's warning that analysis is often treated '…as referring only to quantitative tools generally and operations research and systems analysis more specifically. However, it is important to bear in mind the term's actual definition…' common defence usage loses sight '…of the purpose that actually defines the original term. Gamers supporting DoD [Department of Defense] have too often adopted this usage, referring to games as an "art" that is distinct from "analysis" or even explicitly arguing that games are not analysis even when they are conducted to enhance our understanding of policy. In doing so, they are ceding ground to researchers who have overly constrained the meaning of science and analysis'.[18] This links to the UK MoD's definition of the term 'Operational Analysis', as often the scientific methods to which it refers are implicitly considered to be quantitative. This view is one that we believe undersells the ability of games to provide analytical and valid insights into national security problems using a broad range of qualitative and quantitative methods. However, we also recognise that practitioners need to remain careful not overstate the utility of gaming as a method. Games are a useful method of identifying issues and proposing solutions, but they will not explore the entirety of any proposed problem.

2.3.1 Tenets of an Analytical Game

Having defined an analytical game we have also generated a number of underlying tenets which we believe need to be followed if the game is to be considered analytical in following a coherent process of logical reasoning, and these will be elaborated on throughout this report in sections 3.3 - 3.5.

1. **Employment of Analytical Methods to Generate Insights:** Insights from an analytical game should be to some extent the product of appropriate qualitative

[18] Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, 2020, pp. 171-172.

or quantitative analytical methods that have been applied to the data that was captured.

2. **Verification and Validation of the Game Construct:** The game's construct must be subject to a process of verification and validation to ensure it is fit for purpose, provides an accurate and appropriate representation of the real world from the perspective of its intended use, and that choices made during the design process are transparent.

3. **A Data Capture Plan:** An analytical game requires an appropriate metric collection plan which explicitly identifies what outputs from the game construct will be captured and measured, and identifies the appropriate methods to collect them.

4. **Appropriate Data Capture:** Based on the plan appropriate data must be captured during the game's execution to provide analysts with a proper understanding of what transpired in the game.

5. **Meaningful Post-game Analysis:** Meaningful post-game analysis will be based on insights drawn from a comprehensive understanding of both 'what happened' and 'why it happened'[19] as a result of the employment of analytical methods to the data captured.

6. **Generating Novel Insights:** Post-game analysis of the game should produce insights that are not purely the product of the scenario and/or mechanisms that were an inherent part of the game's design.

7. **Generating Additional Questions:** An analytical game should also generate additional questions that will inform further research.

2.3.2 What Is a Valid Insight?

As a generality the output of a game can be divided into observations, insights and lessons (OILs). Longley-Brown defines these as:

1. **Observations.** Observations are captured from sources, whether they be people or things. Observations are the basic building blocks for insights and lessons identified, but they often offer a limited or subjective perspective on their own.

2. **Insights.** Insights are objective conclusions drawn from *post-game analysis*[20] of patterns or groups of observations.

3. **Lessons Identified.** Lessons are insights that have specific potential and actual authorised actions attached. Lessons *identified* are those that substantiate requests for recommended actions to be authorised. Lessons

[19] Our previous experience is that the 'why it happened' element is often much more difficult to capture/understand, but is critical to undertaking credible analysis that goes beyond simplistic narrative observations of the events which took place.
[20] Italicised text is the author's addition.

> *learned* are those that have been accepted into doctrine, operation procedures etc.[21]

In addition to this, and as outlined above, we would argue that insights from an analytical wargame should be the product of appropriate analytical methods that have been applied to the data which was captured to draw conclusions that are both valid and have been validated.

We believe that it is important to stress this point as players often leave games believing that personal *subjective observations* made by themselves or others during gameplay are actually valid *objective insights*. This is problematic as players can erroneously leave an analytical game believing that they have gained knowledge and that it would be appropriate for them to act on this knowledge. We would argue that players should not take any immediate conclusions away from an analytical game. The purpose of their attendance is to generate observations via gameplay which are suitable for post-game analysis, and it is this post-game analysis that will produce insights which can be deemed as valid. By definition the players alone cannot generate insights. They are essentially part of the system under study; their perspectives are necessarily limited by their role within the game, and they do not know the strengths or weaknesses of the game's model, how it has abstracted reality, or how the game team intend to capture and analyse the data. Any/all genuine insights will come from the game team and be disseminated via the final report, as any observations made by the players need to be validated by the employment of appropriate analytical methods during post-game analysis before they can be classified as genuine insights.

This leads directly on the question of what classifies as 'appropriate analytical methods'. We take a very broad view on this topic. Appropriate methods could encompass anything within the entire range of scientific qualitative or quantitative analytical toolsets that could be applied to data captured during a game. Section 3.6 covers this in detail, but what is important at this juncture is to highlight that at some point prior to the game's execution the design team will have identified these methods and they will have to some extent informed the game's design in order to ensure that the appropriate data will be generated to allow for their proper employment.

2.4 Developing Alternate Analytical Gaming Archetypes

Given that we have suggested some alternative definitions for some of the terminology Bartels is using we also propose slightly altering and adapting her archetypes, the new definitions of which we will employ during the rest of this report:

[21] Longley-Brown, *Successful Professional Wargames: A Practitioner's Handbook,* The History of Wargaming Project, p.138.

		Goal of Game	
		Creating Knowledge (Analytical Methods)	**Conveying Knowledge (Teaching Methods)**
Problem Type	**Unstructured/ Wicked Problem**	Discovery Games	Educational Games
	Structured Problem	Experimental Games	Training Games

Table 2 – Adapted Professional Game Archetypes

We would define a **Creating Knowledge** game as:

> *A game that is undertaken as part of an analytical process whose purpose is gaining insights into a problem space.*

Games that create knowledge are designed to generate novel insights from an analysis of events; insights flow outwards from the game and the participants to the game team, and then to stakeholders via some sort of feedback mechanism. This is in contrast to games designed for **Conveying Knowledge**, where insights that are already known to the game design team flow inwards to the participants, as the game design team employ the game as a teaching method to impart these insights.

In line with our view of an analytical spectrum, we believe that games designed to create knowledge in different problem spaces do so through the employment of a broad range of analytical methods – from potentially highly qualitative methods required to generate insights in Discovery Games, through to potentially highly quantitative methods required to generate insights from what we now term 'Experimental Games'.

We would define a **Discovery Game** as:

> *A game that is undertaken as part of an analytical process whose purpose is gaining insights into an unstructured problem space.*

Given the unstructured nature of the problem space, Discovery Games lend themselves to inductive and abductive methods of logical reasoning as the most appropriate methods of drawing insights.

Inductive reasoning moves from specific observations to broader generalizations and theories, and is often called a bottom up approach. 'In this process of reasoning, general assertions are made based on specific pieces of evidence. Scientists use inductive reasoning to create theories and hypotheses'.[22]

Inductive Method (Qualitative Paradigm)

[22] Lumen Learning, "*Reasoning and Inference*", https://courses.lumenlearning.com/boundless-psychology/chapter/reasoning-and-inference/#:~:text=Abduction&text=The%20difference%20between%20abductive%20reasoning,seeks%20to%20determine%20general%20rules.

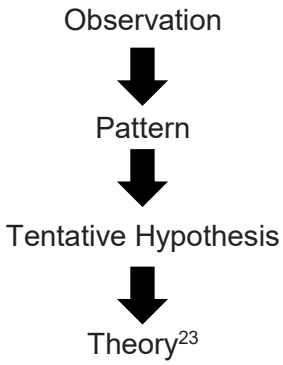


Figure 1 – The Inductive Method

Abductive reasoning, otherwise known as 'inference to the best explanation', is a form of reasoning that 'starts with an observation or set of observations then seeks to find the simplest and most likely explanation'.[24] 'Essentially, this type of reasoning involves making educated guesses about the unknowable from observed phenomena. Examples of abductive reasoning include a doctor making a diagnosis based on test results and a jury using evidence to pass judgment on a case: in both scenarios, there is not a 100% guarantee of correctness—just the best guess based on the available evidence. The difference between abductive reasoning and inductive reasoning is a subtle one; both use evidence to form guesses that are likely, but not guaranteed, to be true. However, abductive reasoning looks for cause-and-effect relationships, while induction seeks to determine general rules… in abduction there is an implicit or explicit appeal to explanatory considerations, whereas in induction there is not'.[25]

The majority of the national security games Dstl has run thus far have primarily used abductive reasoning to draw conclusions, and by extension these games also tend to follow a qualitative paradigm.

We would define an **Experimental Game** as:

> *A game that is undertaken as part of an analytical process whose purpose is gaining insights into a tightly bounded and structured problem.*

An 'Experimental Game' is a game that has been designed to employ analytical methods to consider possible outcomes in bounded and structured settings. A hypothesis to test or well-defined research question(s) to answer are essential to

[23] Adapted from "*Deductive and Inductive methods of logical reasoning*", 2012, https://bhaugolikgyan.wordpress.com/2012/12/11/deductive-and-inductive-methods-of-logical-reasoning/
[24] Longley-Brown, *Successful Professional Wargames: A Practitioner's Handbook,* p.136
[25] Stanford Encyclopaedia of Philosophy, "*Abduction*", 2017, https://plato.stanford.edu/entries/abduction/
Lumen Learning, "*Reasoning and Inference*", https://courses.lumenlearning.com/boundless-psychology/chapter/reasoning-and-inference/#:~:text=Abduction&text=The%20difference%20between%20abductive%20reasoning,seeks%20to%20determine%20general%20rules.

structure the problem, and what constitutes a 'well-defined research question(s)' for an experimental game is addressed in section 2.5.2.4. Experimental Games seek to test more specific hypotheses and assess the possible outcomes of particular actions. By applying analytical methods to more bounded and structured problems, properly executed and applied experimental games can provide a detailed understanding of a well-defined problem. We do not intend for this term to have a substantively different meaning from the definition of 'Analytical Games' as Bartels defined them, but we believe that this captures the same overarching meaning whilst also creating a clearer taxonomy by avoiding the repeated use of the word 'analytical' with different meanings in different contexts. Given the structured nature of the problem space, experimental games lend themselves to more deductive methods of logical reasoning as the most appropriate methods of drawing insights. Deductive reasoning works from the more general to the more specific, and is referred to as a top down approach. By extension these games also tend to follow a quantitative paradigm.

Deductive Method (Quantitative Paradigm)

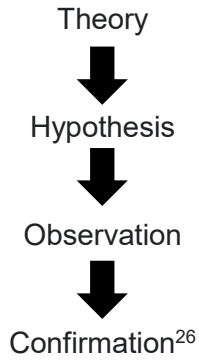


Figure 2 - Deductive Method

[26] Adapted from "*Deductive and Inductive methods of logical reasoning*", 2012, https://bhaugolikgyan.wordpress.com/2012/12/11/deductive-and-inductive-methods-of-logical-reasoning/

2.5 Types of Creating Knowledge Games

2.5.1 Elizabeth Bartels's Archetypes

In another of her works Bartels also puts forward a number of different types of national security games[27]:

	Develop an Understanding of the Problem	Develop Strategies to address the problem
Early research to inform research team and sponsor	**System Exploration**  Elicit and synthesise mental models of the problem	**Innovation**  Develop new decision options that break from the status quo
Mature research to influence external stakeholder in a decision process	**Alternative Conditions**  Detect similarities and differences in decisionmaking based on different starting conditions	**Evaluation**  Judge the outcomes of player decisions based on a normative standard

Table 3 – Bartels's National Security Game Archetypes

These archetypes are linked to the maturity of the research and the overall research goal. Bartels defines each of these archetypes in more detail:

- **System exploration** games: 'try to build out an understanding of a particular policy problem from a range of perspectives... [they] elicit and synthesize how players understand a problem in order to develop a better model of the policy issues, opportunities and constraints. These games are common in early stages of research, intended to form a foundational understanding of the problem system for later studies and analysis'.[28] 'The output from successful system exploration games is a representation or model of the problem that

[27] Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, p.65, 173
[28] *Ibid*., p.76

combines insight from players with research performed by the design team to improve the sponsor's understanding of the nature of the problem…they can be used as a means of developing hypotheses about the nature of the policy problem that can then be further examined using other means'.[29]

- **Innovation** games are games designed 'to generate innovative solutions to policy problems. These games are designed to produce candidate solutions which can then be subjected to further research and analysis, and thus innovation games tend to occur fairly early in research processes… the goal is to generate candidate ideas that make the most sense in the game but will need to be tested out before one can be confident how they will transfer to the real world…These games build a model of the world that relaxes constraints in the hope that doing so might enable new approaches to problem solving. In this way, they share similarities with hypothesis generation and brainstorming activities'.[30]

- **Alternative conditions** games '… seek to understand how a key factor shapes decision making processes and choices... [They are] designed to produce information that helps researchers, sponsors, and consumers better understand the nature of problems by highlighting the impact of alternative conditions on decision making…they tend to come after initial research framing the policy problem; perhaps through a systems exploration game, perhaps through another form of research'.[31] 'Successful games of this type produce an understanding of the influence of varying conditions on either the decision making process or the eventual decisions'.[32] These games are often well bounded to isolate the impact of a particular variable.

- **Evaluation** games 'evaluate policies and strategies… [they] provide information about the outcomes of a proposed solution, whether concept, capability, strategy, or plan, with enough fidelity that the plan can be judged'.[33] 'Because the game must project plausible outcomes in order to enable evaluation of the results of decisions, it must contain a fairly well developed theory of causality that allows the game staff to project different counterfactual outcomes based on player actions. The desired outcome of these games is an assessment of the potential gains and losses from following a course of action'.[34]

Whilst we do not disagree with any of the content of Bartels's archetypes we believe that there is an area between 'early research' and 'mature research' which we feel the current archetypes do not adequately address, despite the fact that she fully acknowledges that the archetypes are a spectrum and not mutually exclusive alternatives.[35] Whilst Bartels directly states that she is not intending to ignore the middle ground we feel that it needs to be explicitly acknowledged and described in the archetypes for them to have strong practical utility. This is particularly important for our research as we also consider a significant number of our national security

[29] *Ibid*., p.62, p.76
[30] *Ibid*., p.63, 114, 173
[31] *Ibid*., p.97, 173
[32] *Ibid*., p.62
[33] *Ibid*., p.125, 173
[34] *Ibid*., p.63
[35] *Ibid*., p.63

games and the problems we analyse to occupy this middle ground – examples and advice relating to the corners of the spectrum are less useful when most of the games we run are focussed on the middle.

We would contend that at the intermediate point in the research there is a clear requirement to refine understanding/strategies that were generated during the early phases, and to refine them to a fidelity that would be amenable to more mature experimental-type approaches. Whilst this technically is touched on as part of her definitions at the early research stage we would argue that there is a substantive difference that needs to be made explicit between games designed to promote innovative thinking and generate novel ideas in comparison to those designed to take such ideas and refine them into a comprehensive and detailed understanding of the problem or 'theory of success'.[36] We will therefore now propose an alternative expanded version of Bartels's archetypes, which we will then employ throughout the rest of this paper.

[36] A "theory of success" is a causal argument [or set of arguments] about what sets of actions are likely to produce the desired result in a specific conflict. – definition adapted from Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, p.68

2.5.2 The Authors' Proposition for an Typological Framework

The overview of our expanded version of Bartels's archetypes is shown below:

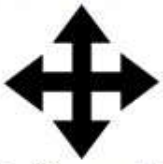
	Develop an Understanding of the System	Develop Strategies to address the problem
Early research using immature inputs which informs the research team and sponsor *(Discovery Games)*	**1) System Exploration**  Characterise and generate an initial model of the system by eliciting and synthesising current understanding	**2) Strategy Innovation**  Develop an initial 'theory of success' by generating new decision options that break from the status quo
Intermediate research using maturing inputs which evolves the understanding of the research team and sponsor *(Discovery Games)*	**3) System Refinement**  Generate a more detailed understanding of the system to refine the proposed model and increase confidence that it is a reasonable representation of reality	**4) Strategy Refinement**  Generate a detailed understanding of the strategy with a view to refining, and increasing the confidence in the efficacy of the 'theory of success'
Focused research using mature inputs which informs external stakeholder(s) in a decision process *(Experimental Games)*	**5) System Conditions**  Detect similarities and differences in decisionmaking based on different starting conditions	**6) Strategy Evaluation**  Detailed implementation of the 'theory of success' to judge the outcomes of player decisions based on a normative standard

Table 4 – The Authors' Expanded National Security Game Typological Framework

Bartels's previous archetypes can be mapped onto numbers 1-2 and 5-6 of our typologies, and they sit in the early and mature input stages. Rather than the left column referring to research maturity – as in the original archetypes – we have altered this to refer to input maturity. The reason for this change is that we believe input maturity to be more explicitly definable in a way that clearly differentiates between the three rows. We have also made minor modifications to the titles and definitions of her original archetypes, with the intent that these changes will help with coherence and clarity in relation to the six typologies we now propose. We do not feel

that our changes substantively impact on or invalidate the more detailed definitions put forward by Bartels as described above, as they are only minor expansions.

Our changes are designed to make explicit firstly, that the left column pertains to our understanding of the system/problem-space as a whole, whereas the right column pertains to the development and testing of a theory of success – i.e. the strategy to address the problem defined in the left column, and secondly, that there are three distinct levels of input, maturity and output which each require different types of games to address. From this point on we will refer to the six new titles we have put forward, rather than those used by Bartels.

Each typology of our new framework will be defined in detail below, and within each definition we will state the expected outputs. We recognise that within each typology there exists a spectrum of different games that could be played which would almost certainly provide a range of different outputs. Rather than attempting to categorise the entire possible spectrum of outputs, our definitions refer specifically to outputs which would represent the culmination of all efforts within each typology – we are essentially defining the outputs as those that are furthest along the spectrum. These outputs could effectively be considered as the inputs for the next typology along in the framework, forming a part of a cycle of research.

Numbers 1 and 2 – renamed System Exploration and Strategy Innovation – sit within the early research and immature inputs row of the table and are usually initial attempts to characterise a system or come up with ideas for effective strategies within it. The outputs of these games are often insights and characterisations of problems, and/or an initial proposal for a solution which requires subsequent refinement.

- **System Exploration** games are designed to elicit participants' perspectives to establish a foundational characterisation of a problem. The output of such games will be an initial model of a system and broad insights relating to its salience, potential boundaries and important aspects. We contend that some degree of characterisation and understanding of the system is necessary before strategies for operating within it can be explored.
- **Strategy Innovation** games are designed to encourage innovative thinking about potential actions to take. They are intended to produce candidate solutions, or 'theories of success' that would be refined and tested further as understanding of the problem, and the proposed solutions, matures. A defining feature of Strategy Innovation games is that players are tasked with coming up with a strategy in the game itself; however, games in this archetype can range from those which output very initial 'blue sky' thinking about broad approaches through to those which help to develop a detailed and coherent 'theory of success'. Such games will almost always require a broad characterisation of the system as an input. Games whose outputs are intended to help produce detailed strategies might also require broad strategies or guiding strategic principles[37] as inputs.

Our additions to Bartels's archetypes – a second row comprising of numbers 3) System Refinement and 4) Strategy Refinement – serve to address the limitations that we identified previously. Games in this row are used to improve models of the

[37] These might include, amongst other things, descriptions of factors deemed to be particularly relevant to a strategy, key capabilities a strategy is expected to exploit, or key perceptions, vulnerabilities or behaviours a strategy should seek to target.

problem and refine strategies to the point where they could be subjected to focused and detailed experimentation and evaluation. Such games require a working model of the problem or a candidate strategy, or strategies, as inputs to the game. These could be derived from games in the first row or other sources. The intent of these games is to refine these inputs by subjecting them to challenge, with the understanding that if problems or weaknesses are revealed the model/strategy is still subject to change and iteration. These games – often in conjunction with other research – challenge the model/'theory of success' that was proposed in the early stages of research to ensure that the key underlying assumptions hold together. Game players are provided with a model or strategy of sufficient maturity that they are able to focus more fully on operating within a system or implementing a strategy, rather than initially characterising a system or inventing strategies.

- **System Refinement** games build on the outputs of early research with the intention of generating a more detailed model of the problem system. Post-game analysis of game outputs will seek to build towards a model that the designers consider to be a reasonable representation of reality. The model generated as a result of the game, and post-game analysis, will be suitable for potential employment in further areas of research such as forming a baseline for system conditions analysis/games and as a context for detailed strategy development. The model should also be verified and validated as far as is practicable for each intended use.[38] To enable these outputs, system refinement games will require a pre-existing characterisation of the system as an input.
- **Strategy Refinement** games build on the outputs of early research with the intention of generating a comprehensive and fully refined 'theory of success' that has been subject to challenge, thereby providing reasonable levels of confidence in its efficacy. The focus of these types of games is on iterative improvements to a strategy through adversarial challenge. Players will be tasked with implementing a well-defined strategy rather than developing one. The outputs of such games will range from identification of areas where strategy improvements are required, through to a strategy or 'theory of success' which is sufficiently mature and of a level of detail that it is suitable for exacting testing and evaluation.

The games that sit in the focused research and mature inputs research row – number 5) System Conditions and 6) Strategy Evaluation – seek to produce more robust insights relating to mature systems models and fully formed/refined strategies. Games in this row are more experimental in nature. To produce rigorous outputs, they tend to focus on testing specific hypotheses or evaluating bounded aspects of a strategy.[39] These games are not designed to make on-the-fly improvements to

[38] The Aqua Book defines verification and validation as follows, 'Analytical quality assurance is more than checking that the analysis is error-free and satisfies its specification (verification). It must also include checks that the analysis is appropriate, i.e. fit for the purpose for which it is being used (validation)'. HM Treasury, *The Aqua Book: Guidance on Producing Quality Analysis for Government*, p.6.

[39] All six boxes of the framework relate to wicked unbounded problems that are complicated in terms of their inputs and outputs. However, to do more focused research into specific areas will be artificially imposed by the game designers in order to allow the problem to be studied. Ideally a system refinement game or games should allow the designers to make an informed decision as to where these boundaries should lie.

systems models or refinements to strategies. Instead, they take more detailed models and strategies as inputs and seek to produce outputs relating to the impact of particular variables on a model, or evaluate the effectiveness of particular aspects of a strategy.

- **System Conditions** games seek to understand how a key factor shapes decision-making processes and choices. They explore the impact of different starting conditions on the system and the choices faced by decision-makers within it. These games are often heavily bounded[40] to isolate the impact of a particular variable.
- **Strategy Evaluation** games seek to evaluate policies and strategies. They provide information about the outcomes of a proposed solution with sufficient fidelity to allow the utility of a plan to be judged.

We would contend that our 6 typologies of game outlined in Table 4 all sit within the two different types of Creating Knowledge games outlined in our analytical gaming framework in Table 3 (copied below for ease of reference):

		Goal of Game	
		Creating Knowledge (Analytical Methods)	**Conveying Knowledge (Teaching Methods)**
Problem Type	**Unstructured/ Wicked Problem**	Discovery Games	Educational Games
	Structured Problem	Experimental Games	Training Games

Table 5 – Adapted Professional Game Categories Copy

We would argue that numbers 1-4 produce broad insights into unstructured/wicked problems and thus fit the definition of Discovery Games. Strategy Refinement games can be employed as an integral part of the analytical process that produces a mature theory of success; we would contend that at this point the theory of success is still subject to change/modification(s), and that the goal of running a Strategy Development game is to produce broad insights about the theory, rather than testing very specific hypotheses relating to its implementation. Only when the research is mature and the problem has been structured is it then possible to run bounded Experimental Games such as System Conditions and Strategy Evaluation games (numbers 5 and 6). These games can test hypotheses and answer specific questions relating to alternative starting conditions and the detailed technical implementation of an already mature theory of success.

Our experiences executing previous national security games highlight the issue that customers frequently wish to obtain outputs of a quality that could only come from an

[40] Given the complexity of the systems in question in some cases these boundaries will be artificially imposed by the game designers for the purposes of creating a workable game framework. Ideally these boundaries would also be informed by the outputs of a previous system refinement game.

experimental game without understanding either the sort of game that would be required to generate such outputs or the quality of inputs in relation to both understanding of the problem and the theory of success that would be required to execute such a game.

2.5.2.1 Typology Objectives

Because of the different inputs, outputs and methods employed within each of the six game types outlined above, in general it is highly recommended to ensure all game objectives occupy just one box within the typological framework. This is because objectives that occupy multiple parts of the framework will tend to create tensions between game design choices that are ideally suited to each. At the very least, this risks creating a very complex game. At the worst, there is a danger that the use of game mechanics which are suited to one set of objectives contradicts or undermines another set of objectives. For example, it is very unlikely that a game designed to generate strategies will also be able to test them. This is because innovation games need to encourage creativity from the players and be open to a vast range of possible player actions. By contrast, evaluation games are built with a particular form of strategy in mind, and only those factors and actions which are directly relevant to its implementation will be included in the design, to avoid unduly complicating the game or shifting focus away from its objectives.

Even within single typologies of the framework, care must be taken to ensure that objectives are not too wide-ranging and unfocused. For example, objectives which seek to develop strategic and operational aspects of a plan in the same game risk over-complicating it, even though they both occupy the 'Strategy Refinement' typology. This is because the players, adjudication expertise and data inputs that are required for each level are different, and the interdependence of strategic and operational factors risks creating a complex feedback loop between players at each level. Ideally, one of these levels of planning should become a fixed assumption or an independent variable which is provided to players as an input. If both levels of planning are to be the subject of player activity, then other objectives should be de-prioritised to ensure the game design is focused on producing high quality outputs in a limited number of areas, rather than low quality outputs in a larger number of areas.

Where such narrowing-down of objectives is not possible or desirable, particular objectives should be nominated as primary objectives, whilst others should be considered secondary. Only primary objectives will drive the game design, while achievement of secondary objectives will be sought to the extent that doing so does not undermine achievement of the primary objectives.

2.5.2.2 Verification and Validation (V&V)

In order to define these terms we have looked to two sources, HMG's Aqua Book and Graham Longley-Brown's *Successful Professional Wargames: A Practitioner's Guide*.

- **Verification** relates to *building the model right*, and is defined by HMG's Aqua Book as 'checking that the analysis [or model] is error-free and satisfies its specification'.[41]

- **Validation** relates to *building the right model*, and is defined by the Aqua Book as checking '…that the analysis [or model] is appropriate, i.e. fit for the purpose for which it is being used'.[42] Validation asks how well the model that has been built is true to reality.

Verification is generally an easier process, as the game designer should have a well-defined conceptual description and specification as a point of comparison. Validation is a more difficult process, especially in relation to a game models that occupy the wicked problem spaces often found in national security games. In particular, elements of validation pose a difficult problem; specifically, understanding whether a model is true to reality, as this implies that we can achieve a comprehensive understanding of a wicked environment in order to compare it to a model. We argue that no model is actually true to reality, and that the model needs to be 'fit for purpose', i.e. that it is accurate to the extent that it helps the game team to undertake the post-game analysis of the relevant area of interest. The model can and should be bounded, and does not need to simulate all of the factors that are relevant to the task at hand as long as it is a reasonable and useful representation and we are fully aware of its limitations; as the aphorism goes, 'All models are wrong, but some are useful'[43].

Game designers therefore cannot definitively say that in all circumstances the model is accurate and realistic, but what they should be able to say is that the model's representation of reality is fit for the purpose that it is being used for.

While the authors acknowledge that it is not possible to V&V a model of a wicked problem to the same extent/level of robustness of a model which is driven by substantial amounts of empirical data, we would argue that it is still possible to perform a reasonable degree of V&V in line with the processes outlined in Chapter 5 of the Aqua Book.

2.5.2.3 Discovery Games – What Is An Appropriate Objective Or Question?

Discovery Games as a group generally require an objective/set of objectives to drive development, but it is acceptable, especially at the early stages of research, for this not to be a specific question that needs to be answered. It is never suitable for the purpose to be a broad statement; for example, "create a Deterrence game" is a poorly worded purpose. Within Discovery Games, System Exploration and Strategy Innovation games are expected to occur early in the research, and are intended to be exploratory. It is therefore expected that these types of games will tend to have the least overall bounding in relation to their associated objectives, hypotheses or questions. Example questions and sub-questions could include:

- What is the nature of the problem?

[41] HM Treasury, *The Aqua Book: Guidance on Producing Quality Analysis for Government*, p.6.
[42] Ibid.
[43] Box, Hunter, Hunter, *Statistics for Experimenters: Design, Innovation, and Discovery*, p.440.

    - What vulnerabilities do we have in domain [X]?
    - What threat might [adversary X] pose in context [Y]?
    - What might the implications be of problem [X] for us?
- What should we do about problem [X]?
    - Does this problem require us to respond?
    - What should our objectives be in deciding how to respond?
    - What should our strategy/policy/plan be to address this problem?

Since System Refinement and Strategy Refinement games are expected to produce more robust outputs than Exploration and Innovation games, their associated objectives should also be more structured and focused. As they occur later in a research process, it is expected that existing system knowledge or nascent strategies will support the development of more bounded objectives. These might seek to focus on exploring or identifying particular aspects of a wider problem, identifying implications in a particular area of an issue or strategy, or subjecting specific, defined strategies to scrutiny and challenge. Example questions could include:

- How might pre-existing relations between actor [X] and actor [Y] affect escalation dynamics in the run-up to a potential crisis?
- What factors might lead to conflict between [X] and [Y] in situation [Z]?
- What challenges might we encounter as we go about implementing this policy?
- What capabilities do we need to deliver this strategy?

Experimental games sit at the mature research level and therefore require a specifically well-crafted question which tightly bounds the problem.

2.5.2.4 Experimental Games – What Is A Well-Defined Research Question?

The purpose of the game is the main driving force behind all of the game design. When moving into mature research areas in structured problem spaces we will be undertaking Experimental Games, and that Experimental Games by necessity require some degree of bounding usually imposed by a well-defined research question(s).

Creating/identifying an appropriate question to answer is therefore critical to the success of an Experimental Game. System Conditions and Strategy Evaluation games are intended to generate highly valid evidence about the operation of specific factors within a system or the likely impact of particular aspects of a strategy. They follow the scientific method, isolating specific variables for careful study. Because they are designed to produce very focused outputs, they require specifically formulated questions which tightly bound the problem. Since System Conditions and Strategy Evaluation games occur later in the stages of research and strategy formulation, it is expected that sufficient knowledge of the system, or detail in a ‘theory of success’ will exist to allow the development of very focused objectives. System Conditions and Strategy Evaluation questions should address either a specific variable or specific aspect of the system, and should be written in such a way that they could theoretically be given a binary yes/no answer (even though this is extremely unlikely to be the case in practice). If the question is broader than this then it would probably not be considered to constitute an experiment.

Whilst it is possible for a game to answer multiple questions at once it is harder to design a game with multiple primary questions and purposes unless they are complimentary. As already noted, McHugh stated that wargames should be 'pointed' towards analytical or training purposes.[44] But even within games designed to create knowledge, the research questions should be focused so that the game design is not attempting to represent or analyse too many things at once. As with all good scientific experimental design the more independent variables that are required to be present the harder it can be to pinpoint and assess their impact on the dependent variables which are the primary contributory factors to success or failure. In general it is therefore best practice to focus on one primary question whilst identifying some secondary questions that the game could answer. It is important to identify secondary questions as this allows the final game design to be used for other purposes if they have been identified as appropriate.

There two broad groups of questions experimental games should look to answer:

a) Solution Questions

b) Aspect Questions

Dividing questions into categories allows designers to focus on what is important within each.

Solution Questions

Strategy Evaluation games tend to focus on *Solution Questions*, which examine what could be done for a strategy or approach to potentially achieve success. The priority for Solution Questions is in testing options available to the players and exploring a specific aspect of the problem space through a lens of pre-identified potential solutions to problems. These questions focus on finding and testing appropriate approaches, ways, means, capabilities and strategies. A-B testing games fall in this category.[45] As a generality these questions more naturally align themselves with strategy evaluation games. Examples of appropriate Solution Questions include:

- Could [theory of success X] allow the UK to achieve conventional deterrence against [peer adversary Y] in [scenario Z]?
- Would [strategy X] allow the UK to meet all of its Humanitarian Assistance and Disaster Relief (HADR) commitments if the rate of natural disasters increased?

Aspect Questions

Systems Conditions games tend to focus on *Aspect Questions*, which seek to generate a detailed understanding of certain aspects of a problem, examining why issues occur in certain ways. The focus for Aspect Questions is on testing and enhancing one's understanding of how particular factors and issues operate within a well-defined problem space. Examples of appropriate Aspect Questions include:

[44] McHugh, F., Fundamentals of War Gaming, US Naval War College, 3rd Edition, 1966, page 9, quoted in DCDC, 'Wargaming Handbook', 2017, Ministry of Defence

[45] A-B testing is a way to compare two versions of a single variable, typically by testing a subject's response to variant A against variant B, and determining which of the two variants is more effective. A-B Testing games typically compare strategies or capabilities (i.e. Strategy A vs Strategy B).

- Do clear red lines help with deterrence strategies?
- Is escalation necessary for successful coercion?
- Does knowledge of an opponent's internal pressures change an actor's behaviour and risk of miscalculation?

3 How Can We Develop "Creating Knowledge" Games That Are More Analytical?

3.1 Previously Identified Issues

Throughout the games in the VCDS series we have identified a number of issues which we believe could be mitigated, and in doing so this would allow our game designs to be more analytical.

As outlined in 1.1 the requirement for this research is predicated on three issues which will need to be addressed:

a) **Experiential Value:** The primary value of senior level national security games has thus far been from the participant's experiences of decision making and the impact of living with difficult choices taken during gameplay – they identified / discovered new issues, derived fresh insights about the challenges of implementing responses, and tested ideas and perspectives amongst their peers. Whilst useful, it is hard to ensure that participants' self-generated insights and lessons are a reasonable reflection of the game design, and take proper account of the assumptions that underpin the game mechanics, scenario and data used in the game.

b) **Identifying Genuine Insights:** In some cases the 'insights' that players drew from the game were clearly a direct product of the scenario/mechanisms baked into the game design by the designers; this was mainly a cause for concern in relation to insights drawn regarding the understanding of Red actions and behaviours. To elucidate further this point, as game designers we know that the dynamism of a game creates a perception that there is a strong relationship between the decisions taken by players and the responses/feedback generated in the game; however, we also know that this perception can be (and often is) false, and is generated from the participants having an incomplete understanding of the mechanical limitations of the game, as well as the nature of the abstractions/assumptions made during the game design process.[46]

c) **Post-Game Analysis:** Comparatively few genuine insights were documented by the data capture processes from the game and then derived from post-game analysis of gameplay, which can provide deeper, and more robust insights than the initial impressions generated by the players themselves. The recognition of this issue leads onto a requirement to address what sort of genuinely useful analysis can be undertaken on highly qualitative discussion based games directly, and what value would this add over our current more straightforward approach based primarily around writing up the narrative.

Further to these issues we have also identified two other issues relating to the tenets that are required to make an analytical game – section 2.3.1 – which this research will also address:

[46] Looking 'under the hood' shows that the adjudication procedures which generate the game's responses are frequently either largely pre-scripted or based on Subject Matter Expert (SME) judgement/opinion, which is often elicited in short timeframes and in high pressure time-constrained situations.

a) **Appropriate Data Capture:** Previous Discovery Games have not prioritised effective capturing of substantive data within their designs. This was partly a conscious choice on the part of the designers to focus on immersion and the experiential value that could be gained from these games. However, this was also driven by a number of practical problems that stem from: firstly, the inherent difficulty of capturing qualitative data from chaotic interpersonal environments; and secondly, the practical problems posed by capturing such data from senior participants in highly classified environments without technological support and in a short time.

b) **Employment of Analytical Methods and Meaningful Post-Game Analysis:** Data captured has generally been of variable quality, volume and detail, as well as being highly qualitative – it has therefore not been amenable to analysis. This has led to comparatively few genuine insights being derived from post-game analysis of gameplay, meaning that insights have been primarily observational and have not been subject to rigorous scrutiny.

3.2 The Gaming Process – Art and Science

In order to be more analytical in our approach to gaming we must first provide an overview of the gaming process. This will allow us to explicitly identify areas of potential improvement. Graham Longley-Brown, co-author of the MODs *Wargaming Handbook* and author of *Successful Professional Wargames, A Practitioner's Handbook*, provides this outline of the 'wargaming lifecycle', which includes the entirety of the game process from conception, design, through development and execution, and then validating the game post-execution as well as writing a final report that includes Lessons Identified (LI) and suggested refinements.[47]

[47] Longley-Brown, *Successful Professional Wargames: A Practitioner's Handbook*, p.197

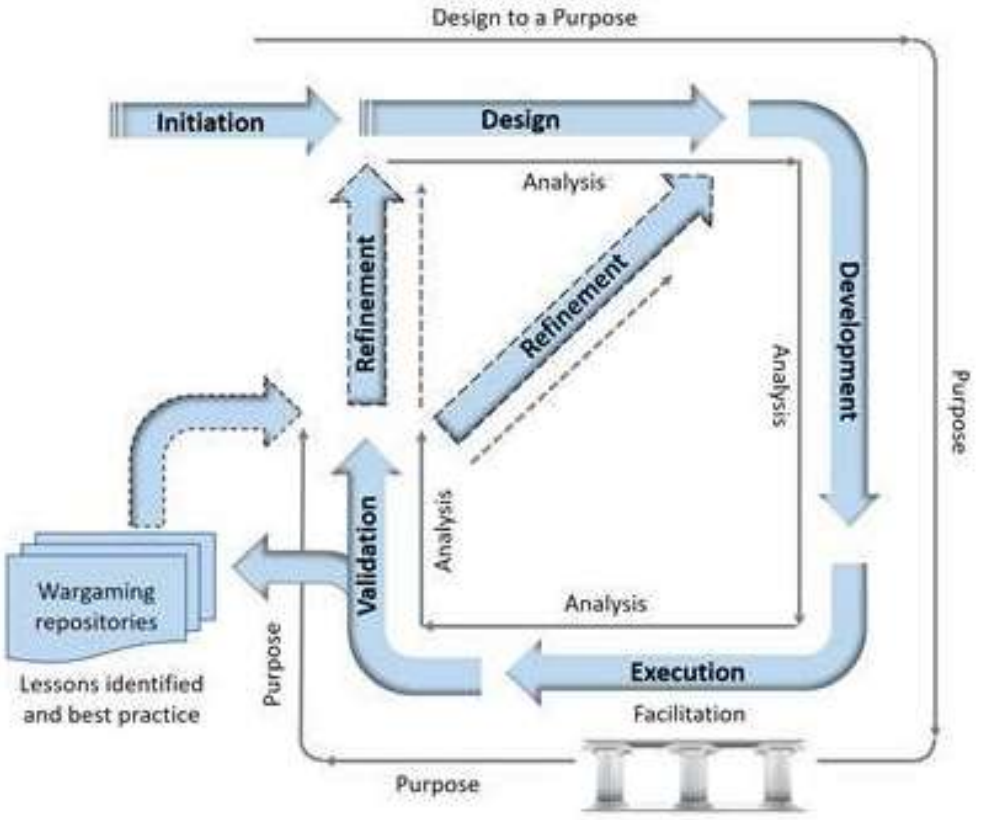


Figure 3 – The Wargame Lifecycle.

Taken from Successful Professional Wargames: A Practitioner's Handbook, Graham Longley-Brown, p.199.[48]

It is important to note that the analysis of the game's output is not included in this lifecycle, which focusses specifically on the game itself as the area of interest. This exclusion recognises that games designed to convey (rather than create) knowledge may not have an analytical requirement, and it remains entirely consistent with our definition of analytical games as those which employ analytical approaches and/or methods to generate insights *as part of an analytical process* – the analysis of the data is part of this wider analytical process.

Amongst practitioners there is a long running debate as to whether games should be considered an art, a science, or both. It is not the intent of the authors to fully address this debate; however, we believe that identifying the contributions that practitioners consider 'art' and 'science' make to wargame design and analysis will allow this research to identify specific areas where the application of different analytical processes would help to generate more meaningful and valid insights from our games.

Proponents of more scientific approaches to wargaming argue that overemphasis of the artistic aspects of games limits the ability to objectively judge their validity and

[48] *Ibid*., p.199

quality[49] and risks producing outputs that are misleading or of little use to sponsors[50]. Nevertheless, even those arguing for more science in game design acknowledge that "truly masterful game design requires creativity and art[51]". Others counter that judging wargames by scientific standards of rigour and objectivity risks failing to appreciate the true value of wargames[52], and understates the importance of the inherently variable, difficult to codify, human element of decision-making in wargames[53].

Longley-Brown argues that:

> *'Wargaming – like war – is as much art as science. I contend more so… wargaming has elements of both: it is an art form overlaid on rational science. The problem is that art is hard. Creating an effective wargame requires craft and intuition... no amount of rational planning will obviate the requirement for the artist/designer to rely at some (probably frequently) point on instinct… it takes an inveterate wargamer with multiple skill-sets plus something indefinable to design and deliver a true story-living experience whilst resolving the inevitable glitches along the way'.*[54]

We believe that generating more meaningful and valid insights from our games can be accomplished firstly by suggesting improvements to the processes and methods we already employ in areas that are considered to be 'the science of gaming'. We would argue that as a generality these areas encompass:

- The inputs which underlie the game's model of reality, 'doing the background research, sourcing the data, understanding the underlying theories for the phenomena you need to represent, and understanding the processes and the causal mechanisms… [alongside] credible models (games, mechanics, sub-models, etc.) to explore the issues under consideration… [and] determining how credible these models are'[55].
- The development of the game's Data Collection and Management Plan (DCMP) – a document which should define all aspects of the questions to be answered, the variables and measures which will be employed, how data will be collected, and what analytical methods will be employed to derive insights.[56]

---

[49] Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, pp.39-40
[50] Compton, "The Obstacles on the Road to Better Analytical Wargaming," *War on the Rocks*, 2019, https://warontherocks.com/2019/10/the-obstacles-on-the-road-to-better-analytical-wargaming/
[51] Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, p.36
[52] Perla, Ewell, Ma, Peachey, Sepinsky, and Tripsas, "Rolling the Iron Dice: From Analytical Wargaming to the Cycle of Research," *War on the Rocks,* 2019, https://warontherocks.com/2019/10/rolling-the-iron-dice-from-analytical-wargaming-to-the-cycle-of-research/
[53] McGrady, "Getting the Story Right about Wargaming," *War on the Rocks,* 2019, https://warontherocks.com/2019/11/getting-the-story-right-about-wargaming/
[54] Longley-Brown, *Successful Professional Wargames: A Practitioner's Handbook*, p.32
[55] Pearce, Email Correspondence with Author, 23 Apr 2020
[56] Longley-Brown, *Successful Professional Wargames: A Practitioner's Handbook*, pp.136-7

- Undertaking the post-game analysis of the data collected to generate insights.
- Data capture.

Secondly, we would also argue for the expanded usage analytical methods into an area of game design and development that most practitioners would currently consider to fall under the auspices of 'the art of gaming'; this being the ability of the game design team to abstract the game's underlying inputs into credible mechanics, creating a functional game which works effectively during execution whilst serving the analytical requirements put forward in the DCMP. We believe this has the potential to lead to games that have more transparent and interrogable designs.

3.3 Analytical Game Design, Mechanics and Adjudication

This section looks at a number of core concepts and mechanics in National Security Games that affect our analytical approaches, including the mechanics we choose to use and the adjudication techniques we choose to employ. A key consideration throughout is that the main focus for a good analytical game is **what is the purpose of the game?**

The purpose of a game should be the principal driver for the approaches taken. No one type of game is a panacea for every purpose and often there may be multiple ways of designing a game to answer a purpose. This is a complex task, and as Bartels puts it, 'Existing texts on game design stress the importance of linking design to purpose but offer very little advice on how to achieve this goal. The most often-cited handbooks on the design of games stress the importance of linking the choice of design elements to the purpose of the game, since a "wargame's objectives should be the principal drivers of its entire structure." However, when it comes to *how* to make the linkage, these texts are largely silent'.[57] There are advantages and disadvantages to every approach, and it is important for designers to be aware of these when they make their design decisions.

3.3.1 Free Play Vs Codified Analytical Games

Free Play Analytical Games

Discovery Games usually require players to have large amounts of freedom to make choices and take actions from a broad spectrum of capabilities with varying outcomes; this is known as Free Play. Seminar games are the primary form of Free Play games we engage with. In these games, players are free to take any actions they can think of, whilst being constrained by the scenario and the realities of the actor(s) they are playing. Seminar games allow players to take *approaches* rather than just actions on each turn, by packaging together a range of actions.

However, seminar games rely firstly on players having the appropriate level of subject matter expertise to take plausible actions, and secondly, on adjudicators who are confident and knowledgeable enough to adjudicate multifaceted multi-level interactions in complex problem spaces. The difficulties of adjudication are exacerbated by the fact that the players' freedom to generate approaches during the

[57] Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, p.18

game means that there is no real advanced planning that can be undertaken ahead of time.

Additionally, to adjudicate such interactions properly can be a slow process, with adjudication cells requiring hours to understand/interpret player actions and generate appropriate responses which show a nuanced view of how “the world” has developed each turn. Quick adjudication is possible, but it effectively relies on the adjudicator using their own expertise to craft responses to the approaches taken – this can leave the outputs prone to bias as they are formulated by adjudicator expertise, experience and assumptions, which may turn out to be incorrect. This is why we usually employ an adjudication cell rather than a single adjudicator to ameliorate the issues of individual bias. However, even this mitigation cannot account for group bias, which can lead to unrealistic adjudication decisions that affect the game’s output. It should be noted that as long as experienced SME adjudicators are used it is unlikely that an entirely unrealistic outcome will be adjudicated, but this in itself can be a problem, as often this will lead to an avoidance of unlikely circumstances or plausible ‘edge case results’ as being unrealistic. Such adjudication also takes place entirely within the mind of the adjudicator, leading to a lack of transparency as to why adjudication decisions were made. Such a method is also not auditable post-game – and is therefore not amenable to analysis – and is likely to remove richness from the game insights and greatly exacerbate adjudicator bias. This method of adjudication is often used in games with senior participants as their time is limited and despite its flaws we do not have many other viable methods of adjudication, with Set Narrative/Scripted games being the other method we frequently employ.

Set Narrative/Scripted games provide a common, and sometimes necessary, solution for the problems associated with adjudication in free play games. A set narrative is created for the game during its development that progresses through pre-scripted events designated by the game designers (these sometimes include a small number of potential branch points to enable some degree of dynamics in response to player actions). This approach ensures that analytical aspects of the game are touched upon and also prevents the game from bogging down during adjudication. However, it also partially removes the adversarial nature of the game, as player actions have only a limited ability to change the narrative of the game. These games are still very useful for insight generation, and a Red Team or Cell is still important in helping to generate these insights. However, the lack of adversarial action means that they are weak in certain analytical areas, such as if a more experimental A-B Testing game is being conducted.[58] Furthermore, there is the risk that by railroading the game in a certain direction the designers are constraining the insight space in such a way that the Set Narrative/Script could be driving the insights more than the players.

In a free play game players are by definition free to take any actions they consider to be plausible. As previously noted, this is a strength and a weakness of the format, as by extension free play games for creating knowledge rely on SME players acting realistically/plausibly when undertaking their roles. This is however much more easily said than done, as in many national security games roles are often high level amalgamations/abstractions of many roles and/or organisations (such as UK Government, Public Opinion, North Atlantic Treaty Organisation [NATO], etc.). The lower the level of a game the more consistently reliable SME decisions are likely to

[58] A-B Testing games are games that compare strategies or capabilities (i.e. Strategy A or Strategy B).

be, as their role in game is more likely to accurately reflect their role in the real world due to a reduced requirement for abstraction. It is therefore much more difficult to represent 'UK Government' with players, even via a team of SME's, as the in-game role is heavily abstracted and represents an amalgamation of lots of real-world roles, decision makers and advisers.

In this context, the main strength of Set Narrative/Scripted games is that the narrative is controlled, and is calibrated by the designer so that its progression and insight derivation is from a realistic baseline, albeit a constrained one. The main advantage of following a more adversarial and free form game is the narrative is unconstrained, meaning that valid insights can be drawn relating to its progression. Free play games therefore have a wider insight base, are more likely to derive more niche and creative options as well as exploring second and third order effects more effectively.

Codified Analytical Games

Codified games are games that follow a set of rules with a more rigid structure. Player actions are constrained to a set of pre-determined actions with rules on what each action achieves. Despite actions being constrained, codified games can be very different in terms of complexity, length and breadth of subject. Codified games can be simple and quick to play with a single focus, or complex and multifaceted taking hours to complete one turn.

One of the main strengths of Codified Games is that adjudication is hard baked into the rules and player actions are always directly linked to outcomes. Whether the player's desired game end state occurs could be impacted by chance or counter actions taken by the opposition. The implementation of a pre-determined transparent, structured and rules based adjudication method removes the issues of on the spot adjudication and can engender a more adversarial and dynamic game. Players can read and understand the rules and will not have to guess what their actions will achieve.

Codified games also have the advantage of usually being more comparable, as the rules create an environment that constrains player actions to a more structured pattern – the game rules effectively draw system boundaries. This can sometimes, but not always, allow for more quantitative analysis of the game depending on the game design and whether it is appropriate to do so. They also remove some of the previously identified bias from adjudication, as if players are following a well-defined rule set then adjudication cells are not really necessary (although an adjudicator for fringe cases or unforeseen actions should usually be provided). By constraining actions they also reduce the comparative requirement for players to employ expertise and knowledge to take actions that are realistic. Effectively they shift this burden to the game designers to provide the players with a list of actions that are plausible, and this helps to counteract bias which players could potentially bring to the game.

However, to shift this burden, the rules and structure of a Codified Game must be treated to a higher level of scrutiny. These games should be treated more like computer models, with log books and validation/verification testing before use. It is sometimes not necessary to go through the full rigorous process depending on how the output of the game is being used, but in general if this process is not followed then the evidence from the game should not be treated as a higher level of evidence than a free play game.

The design and development of a codified game requires a bounded problem space and a detailed understanding of the variables under study to create the model. Codified games thus lend themselves more to the Experimental Game space rather than the Discovery space, as they require higher level of understanding and evidence to create actions and mechanics. This is often not possible in the Discovery Game space, as the detailed understanding of the inputs and outputs necessary to create such a bounded model may well not exist. This can also make Codified Games very expensive and time consuming to design as the research can be an extensive task and the burden of evidence for them can be extremely high.

By extension, a significant issue with Codified Games is that to create a high quality analytical game the design team must have expert level knowledge of how the subject of the game works (e.g. deterrence, conventional warfare, hybrid warfare etc.). This is because the design team need to know:

- what actions and mechanics should be included to represent the issues at hand;
- how to abstract these interactions into simplified game mechanics which serve as suitable proxies for variables to be measured;
- how these actions and mechanics will interact player with player actions to engender complex re-world interactions that lead to realistic outputs.

This is incredibly difficult at the Strategic and Grand Strategic Levels for a number of reasons, the principal reason being that experts often disagree on interactions and outcomes and there is no real ability to conduct experiments and simulations to aid our understanding. For conventional warfare, even up to Divisional or Corps Level warfare, it is possible to use weapon statistics, vulnerability data, lower level models, exercise data, trials data and historical data to help design current and future focussed codified games. Furthermore, things like conventional warfare are codified in their own right through Doctrine and TTPs (Tactics, Techniques and Procedures) so it is easier to structure these types of games. When it comes to higher level issues or less structured problems like Influence Campaigns, Grand Strategic Games, Hybrid Warfare etc. it's much harder to structure these games correctly without railroading them down certain avenues of exploration.

Whilst a well-constructed Codified game should reduce the adjudication and player bias of more Free Play games it will not entirely remove bias in the round. The designer will be baking actions and outcomes into the rules and will therefore be moving assumptions and biases into the game design process – if the design makes an action strong, then the outcome will be strong during gameplay. This can have a significant impact on the game's outcomes – especially if players are unaware of what assumptions have been made in the construction of the model. Furthermore, the data itself often is loaded with biases and some form of understanding of this bias is necessary to ensure appropriate caveats are recorded. This is why it is important that the factors in the game are properly researched and verified and validated to ensure that this bias is removed or at least acknowledged as assumptions in the game design.[59] This problem is covered in more detail in sections 3.6 – 3.5.

[59] Sometimes all that can be done is acknowledge assumptions in national security games because in some areas there is no consensus on how some interactions work, especially in political theory.

Another weakness of codified games is that they are closer to games of skill than other types of games. As they are based on structured rulesets, turns and moves can be optimised and players can even exploit perceived weaknesses in the ruleset. For this reason an adjudicator, often the game designer, is still required during the game to offer advice to players as well as recognising when a rule is being exploited and/or not working as intended. This is usually not a problem in professional games as the participants are working together for a common purpose even if they are on opposite sides of a game. However, it must be acknowledged that some players may have a natural talent for learning rules and applying them in a gamic environment, and this can potentially skew the results.

In conclusion, Codified Games are generally easier to analyse as their outputs are easier to measure. However, they require a better understanding of the real world interactions that affect the subject matter of the game, and this may not be possible without a huge amount of additional research. Games should be treated like force multipliers to understanding of a subject, they cannot make up for a lack of research and knowledge in a certain area but they can augment understanding in a subject area greatly. It must be noted that customers often ask for Discovery Games in subject areas that are not well researched or understood, when the creation of a codified game would be a complex task and not an effective use of resources. Codified game approaches are therefore better suited to Experimental Games.

Mixed Analytical Games

It is possible to create a game structure that straddles the Free Play vs Codified dynamic. These games have more structure than a total Free Play game but have fewer rules and set actions than a fully codified game. Many Matrix Game[60] designs achieve this and this partially explains the popularity of the method; however there are other mixed designs that can be created which are not Matrix Games.

In general mixed games have a discursive element within them that is adjudicated. How this element is adjudicated is normally the main difference between these types of games. However, an issue with any discursive driven game that players who are more eloquent and better at debating usually have an edge over other players. This can skew results and is something that facilitators and adjudicators need to be cognisant of during games.

Mixed Analytical games are not always desirable for a number of reasons. Sometimes designers will need the ability to work with entirely free actions and

[60] Matrix gaming is a multi-sided, free-form gaming method in which game actions are resolved, and the game narrative emerges, through a structured process of discussion among the players themselves. The key thing to understand about matrix gaming is that it is up to the players and facilitators, through discussion, to determine the likelihood of successful actions and their consequences. Players state an *action* they will undertake; the *effect* that action would have if successful; one or more positive *pro* arguments why their action would be successful—assuming, of course, they are able to muster any reasons for success. Other players are then free to weigh in, offering additional *pro* arguments for success, or *con* counterarguments as to why the action would fail. Two six-sided dice are then rolled to determine outcome. Each strong and credible *pro* in favour of the action generates a +1 dice roll modifier, and each strong and credible *con* against the action generates a -1 dice roll modifier. In this way the arguments made by players have a direct impact on the likelihood of success. Summarised from *Matrix Game Construction Kit: User Guide,* Brynen, Mouat, Fisher, pp. 7-11.

require a Seminar Game or a more Free Play Matrix Game, and while Mixed Analytical Games can draw on the strengths of both, they may also draw on their weaknesses. Sometimes designers will require harder comparisons between different approaches and will need to use a more codified approach. Mixed Analytical Games do however allow for games to link player actions to more solid outcomes with second and third order effects than a Free Play Game does, whilst also permitting more creative approaches to problem solving and insight generation than a Codified Game could offer.

3.3.2 Deterministic Vs Probability Derived Outcomes

A core part of game design is deciding whether adjudication should be entirely deterministic or have some probability applied to outcomes. Deterministic outcomes will usually either come through adjudicated decisions or player consensus, with some codified systems using a deterministic resolution mechanic. Probability derived outcomes (whether using dice or some other methods) can range from set percentage chances in codified rules to adjudicated ranges of probability of outcomes.

Deterministic derived outcomes have the advantage of following the perceived most likely course of events and being auditable to explain their derivation. This is as long as the determinism is based on sound assumptions and judgement. A common technique in Free Play Games and some Mixed Analytical Games is to determine outcomes through the consensus of players and adjudicators, with possible outcomes being agreed upon by the group and then the most likely outcome being identified and chosen. In other games this could be a purely adjudicator decision with no consultation with players (which is sometimes necessary for expediency).

The main disadvantages with this mechanic is that it:

- does little to reward high risk, high reward strategies;
- encourages conservative decision making whist simultaneously allowing players to ignore some fringe issues;
- does not explore some fringe but important outcomes
- does not necessarily generate more realistic outcomes, even if they are more believable to the players.

For example, if over the course of a game ten outcomes were chosen that had a 60% chance of happening each, then the eventual pathway chosen would have an extremely low probability of actually occurring (0.6%), and it is therefore reasonably likely that at least one of those instances would not have happened. The adjudicator or facilitator can step in to choose a less likely outcome, and sometimes this can be done for good analytical reasons depending on the purpose of the game – it is however critical that such interventions are logged and recorded for the purposes of post-game analysis.

In codified games, deterministic mechanics usually consist of mechanics that specify outcomes of combinations of player decisions. Rock, Paper, Scissors is an example of a codified game with deterministic outcomes.

Probability derived outcomes can share a lot of mechanics with deterministic mechanics in deriving the percentage chances of success, the difference being that random number generation is used to determine which outcome occurs. The main advantage of this approach is that it adds an identifiable element of risk and allows players to use high risk high reward strategies whilst highlighting where failure may occur. The major issue is that chance can skew the outputs of a single game if edge case roles of the dice occur. Codified Games tend to use probability derived methods to help add this element of risk but also to highlight good or bad actions in certain circumstances. This method is therefore likely to create more reliable outputs if the game can be executed repeatedly. Probability derived outcomes also mean that players need to make explicit decisions about the level of risk they are willing to accept, Furthermore, there is a certain amount of stigma that is directed towards this method which can affect credibility.[61] Both methods are, however, analytically sound depending on the purpose of the game.

3.3.3 Turn Structure and Time In National Security Games

Turn Structure is a core part of any game's mechanics and dictates the pace and scale of a game. In terms of turn structure for analytical creating knowledge games there are a number of concerns that need to be addressed:

1) How many different actors are present in the game?
2) What sort of actions are expected from the players?
3) Do the turns have to represent set time periods or different time periods each turn?
4) Should players have the same turn structures or asymmetric turn structures?

1) How many different actors are present in the game?

This consideration dictates whether the game can use simultaneous turns or whether it will be necessary to have actors taking turns one after the other. This can affect the type of game that can be used. For example, Matrix Games cannot deal with simultaneous turns in a traditional manner, and will be extremely slow to execute if large numbers of actors were to be represented – this is why they generally represent 4-6 actors.

The more actors that are represented, the more time that will need to be given to diplomacy within the turn structure, as key actors will potentially need to have a large number of discussions. However, whilst key actors may be overburdened there is the potential for 'dead time' – where players are waiting for other players to finish their turns – amongst niche actors. Although this is not really an analytical consideration, trying to ensure that people are engaged in the room as much as possible is good game design practice.

2) What sort of actions are expected from the players?

Is the game looking for high level considerations or lots of low level detail?

[61] This stigma is often attached to the use of dice as a method of random number generation. Whilst dice are a perfectly acceptable method to represent chance in games they can undercut player's perception of the event as a serious enterprise.

Are we looking for counter actions or troubleshooting a particular approach?

Such considerations should be taken into account when making game design choices which may shape the nature of player interactions. These game design choices could include thinking about which actor plays first, whether turns need to have multiple phases to capture different details, and whether players are given the opportunity to provide counteractions to enemy activities or not.

There is no set way to structure the turns of a game, but the structure should be specifically tailored to support the purpose of the game and enable data capture of player discussions and the eventual actions taken.

Designers should think about what data needs capturing, and from that what sort of player interactions will generate this data and how these decisions will be captured? If there are actions that are not being recorded or that do not fit in the turn structure consideration should be given to whether these actions are needed or if the turn structure needs to change to enable them to have more prevalence. These considerations are covered in more detail in section 3.6.

3) Do the turns have to represent set time periods or different time periods each turn?

Deciding how long a turn represents in real time dictates what players need to achieve in their turns and the pace of the game. If game turns represent long periods of time (a year or years) then it is likely that the game requires a phased structure where players can react to other player actions. Shorter time periods may not require this, as actors can simply react to adversaries' actions in their next turn.

It may be the case that a game will need to represent turns more fluidly, with each turn potentially representing different periods of time – the length of which may sometimes be purposely kept fuzzy to allow for player interpretation. In this circumstance the real time periods should not be too dissimilar, for example shifting from a month to a year, as this would skew other parts of the game design. However, not setting the time period in stone allows for appropriate time skips whilst also being able to focus on certain key events or aspects in the game.

If a set time period is chosen then the main consideration which underlies this is the purpose of the game. Drawing on a relatively simple example of Divisional Level Warfare, the Divisional/Brigade Overlay Wargame System (DBOWS) has a 30 minute per turn real time period, as this is roughly the time it takes to reload a key system under study. As the purpose of this game was to test low level military equipment in the Divisional context and this system in question was the longest to reload tactical system available, it made sense for turns to be based around it so that no capability was reloading over multiple turns. This is a harder consideration for National Security games, and a genuine consideration is often the constraint of how many turns the players will actually have time to undertake. Inherently, less time to execute the game means that turns will probably have to represent longer time periods.

4) Should players have the same turn structure or asymmetric turn structures?

Changing turn structures can be a way of helping represent different sorts of actors in games and can help with the issue of some actors being more represented by others. Analytically there is no problem having players that have different turn structures or

rules over others as long as their inclusion is justified via research and is included in the game assumptions.

3.3.4 A Framework for the Game Design Process

There are a number of ways to design a game with an emphasis on analysis. The process outlined below highlights one of the best case scenarios given enough time and resources to design a game properly. This process essentially provides a more detailed breakdown of the design and development steps shown in the game lifecycle in section 3.2. Certain parts of the process are significant enough that they have been selected for more detailed coverage in their own separate sections, and this will be cross referenced where necessary.

Some of these steps are essential. Deriving the purpose of the game and creating an appropriate question to answer should be the first part of any game design. However, there are some steps, like the Pathways Analysis, that are dependent on the type of game that is being created. These have been marked appropriately.

The basic process is:

I. Deriving the Purpose of the Game
II. Subject Research
III. Create a Concept of Analysis
IV. Scenario Derivation
V. Initial Game Design
VI. Red Team Design
VII. Pathways Analysis
VIII. Second Game Design period
IX. Data Collection Techniques
X. Game Validation and Verification

An important part of this process is returning back to the start after completing the next step to see if anything has changed and needs updating. Some of these steps can be conducted concurrently to each other, but this has to be determined on a case by case basis.

I. Deriving the Purpose of the Game

The purpose of the game is the main driving force behind the game's design. The clearer the purpose of the game the more likely designers are to get a good analytical result, and as outlined in section 0 our typological framework for creating knowledge games essentially comprises of six boxes each related to a different purpose for running the game. On this topic, Bartels notes that:

> *The information that we desire the game to produce is, ideally, another way of stating the game's purpose and objective. However, experienced policy gamers frequently note that the purpose and objective of a game are a frequent point of sponsor intervention, leading to vague or cluttered guiding statements. Ideally, this is solved by the designer guiding the sponsor to generate tight, focused statements of intent about what information the game ought to produce. However, in practice, game designers are often forced to*

> *accept unfocused objectives, and opt to develop a more defined scope for the deliverables with the sponsor informally.*[62]

We argue that to shape the game's purpose into something focused and usable to the game design team the purpose should ideally be expressed either as an individual (or set of) objectives, hypotheses, or questions that meet the overarching criteria set forward in sections 2.5.2.3 and 2.5.2.4.

II) Subject Research

This phase is about conducting research into the subject matter of the game to identify factors that influence outcomes of various actions, approaches and strategies. This step is essential to generate the baseline understanding required to abstract the real world into appropriate game mechanics.

Even in a Free Play Game this step will be important as it will also provide the background research with helps with the creation of country profiles and other relevant supporting materials. In a Free Play Game the designer does not necessarily have to be intimate with the Subject Research as long as someone from the research team is available for the game.

If the game is going to use Probability Derived Outcomes, especially in a Codified Game, then this phase is the phase to try to collect appropriate data to help create these tables. It may be that during this research more assumptions are made to integrate historical data the designers consider to be relevant, and in that case these assumptions should be recorded in the Concept of Analysis.

III) Create a Concept of Analysis

After identifying an appropriate purpose, objectives, question to answer or hypothesis to test, it is important to write a Concept of Analysis. It should lay out the point of the study, the methods to be employed and the analysis techniques that will be used.

The Concept of Analysis should be a living document that the designer comes back to and updates as the design process goes on. In the Concept of Analysis, the designer should record the primary question to be answered as well as any secondary questions they believe the game can answer. Many of these secondary questions will be discovered throughout the design process and some will have to be deleted if changes to the design mean they are no longer relevant. In the first iteration of the Concept of Analysis, it is useful to record any desired secondary questions that stakeholders and customers may have liked to be in the game as a reminder.

The data requirements and analytical planning that are central to the Concept of Analysis are discussed in detail in section 6. Assumptions, issues and risks should also be recorded in the Concept of Analysis and kept up to date. The assumptions in particular are important in ensuring that the game is useful beyond its primary purpose, as they will help identify differences in similarly designed games with slightly different assumptions.

The Concept of Analysis is an important document both during the design process and after the game has been finished. The Concept of Analysis should not be

[62] Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, p.61

changed after the game has been delivered for the first time; the iterative process essentially ends after the game design has been completed.

IV) Scenario Derivation

Scenario Creation is in itself an important skill that needs to be executed correctly and in accordance with MOD guidelines. A game designer does not need to create a scenario for every game and should use appropriate MOD approved scenarios if they are available.

However the scenario is derived, knowing the scenario early in the process can help with the design phase for a number of different reasons. These include providing an understanding of which and how many actors need to be represented in the game construct, geographical considerations, and any other factors that may be unique to the scenario. There are game designs that do not have scenarios or represent scenarios in different ways – if this is the case then that should be recorded in the Concept of Analysis.

Once a scenario is picked the designer should go back to the Concept of Analysis and update, as this may add to the secondary questions and data collection plan.

V) Initial Game Design

This is the first attempt at designing the game and includes research into specific game designs and playtesting of different systems. Some of these processes can start before the previous phases have finished or begun, but initial design should probably not end until phases 1-4 have been completed.

By this stage the designer should be widely aware of the problem space, the data requirements, the scenario, the actors that need to be represented and the factors that need to be included. Initial design will allow for early play tests of concepts.

Production of concept materials for the game, including maps and counters should start during this phase to ensure that the materials used are useful.

VI) Red Cell/Team Design

During initial game design, the integration of the Red Cell/Team into the game should be considered particularly with a focus on what role they should have during the game. Integrating some of the Red Cell/Team into the design phase is important for Red Cell/Team buy-in and to ensure that their expertise is included in the game. See section 5 for further discussion of Red Cell/Team design considerations.

Appropriate SMEs need to be identified for the game based on subject matter and adversaries. If needed, training opportunities need to be identified at this point if such SMEs do not exist.

This process is also important in avoiding bias from the Red Cell/Team. The design team and adjudication cell will have a chance to question Red assumptions outside of the game and allow the research team to double check said Assumptions. All Red Assumptions should be noted in the Concept of Analysis.

VII) Pathways Analysis – (Non-essential)

This phase is non-essential as it may not apply to some game designs and is an augmentation to Red Teams and game designs. A Pathways Analysis would be particularly useful for Set Narrative Games for example.

This technique essential creates a series of pathway events that could happen from a single beginning event. Essentially this is workshopping the game beforehand. This pathway can be used as either an actual script for a Set Narrative Game or as a handrail for adjudicators and/or the Red Team in more adversarial games.

By involving the Red Team in this process, the designer has the ability to create some likely responses to Blue actions ahead of time that can be worked with some detail to include in the game if appropriate, alleviating some burden on adjudication and potentially keeping Red to pre-agreed, sensible pathways.

VIII) Second Game Design Period

By this period the designers should have a strong handle on the design of the game and its structure. The Red Team should be integrated and used for play testing during this time.

At this point the Concept of Analysis should be updated with a more firm direction with secondary questions and assumptions being confirmed.

Production of the final game materials should start during this phase.

IX) Data Collection Techniques

This phase should be the final iteration of the data collection techniques being included in the concept of analysis. This phase includes production of any materials, surveys, spreadsheets etc. required for data collection and acquisition of any other recording equipment required for data collection.

Data collection and analysis activity needs to be tested *as well as the game* when running pilot activities, and training of the scribes and recorders should happen ahead of time as these parts of the game are often parts that slow the process down. Getting more accurate real world estimates of how long each turn is likely to take is important in working out timings of the game and may lead to having to streamline the mechanics. Furthermore, ensuring that the scribes are recording the data correctly is an important output of this point.

Section 3.6 below provides a further discussion of appropriate data collection techniques for national security games.

X) Game Validation and Verification

The full process for Validation and Verification is set out in the Aqua Book. However, games do not necessarily have to meet the full model requirements set out in the Aqua Book.

The minimum requirement is play testing and technical review, which provides some level of validation. The Concept of Analysis can be used as the basis of a Log Book. If possible, the system should be tested against historical scenarios and external review of the system is also desirable.

3.4 A More Scientific Approach to Game Design and Development

As identified in section 3.2 there are a number of possible improvements we can make during the design and development phases of the wargaming lifecycle to generate more meaningful and valid insights. None of the ideas put forward in this section are revolutionary, in fact, most are simply the application of best practice in game design which has, for various reasons that this section will explain, not been fully applied in relation to our national security games.

When designing and developing games that involve computer simulations, the Dstl wargaming capability follows a rigid set of procedures to ensure that the model is realistic and fit for purpose. To summarise, this involves a substantial amount of background research into the phenomena that are being studied to understand the underlying processes and causal mechanisms that explain it. This leads to the identification of input data required to generate a credible model, which is then collected so that the model can be constructed and tested to ensure that it is credible, functioning correctly, and producing the required outputs.[63] During this process the assumptions and decisions made by the design team are catalogued so that the limitations of the model can be made explicit. This ensures that when it is employed the game team know what outputs can be considered valid, and which may be artefacts of the aforementioned limitations.

Our national security games do not follow such a rigid process. Computer simulations are designed to model structured problems that can be quantified to a comparatively high degree of accuracy and precision (within the given assumptions, constraints and identified limitations of the model). National security games deal with problems of a different nature that involve:

- strategic actions
- measures short of war
- geopolitical complexity
- numerous actors
- diplomacy and interpersonal interactions
- a high degree of Clausewitzian ‘friction’.[64]

It is currently not possible to create a quantitative cause/effect computer model which fully captures the vast complexity of the systems we wish to study without massive abstraction, and so no such models exist. Furthermore, national security problems are much broader and more changeable than large-scale operational warfare, which is comparative more bounded and generally changes in an evolutionary and incremental way. Given that the character of operational warfare remains comparatively static if the design is done well the models represent the underlying understanding of the nature of combat at the level of abstraction represented. This ultimately means that new situations can be represented through changes *in data*, rather than changes in model. Therefore, once created such models remain valid for extended periods of time and can justify their cost through repeat usage and incremental development and expansion, and there are also an extended number of

[63] Pearce, Email Correspondence with Author, 23 Apr 2020
[64] Rhodes, “Friction”, 2008, http://www.strategybydesign.org/friction

communities of interest relating to operational warfare who have been willing to fund this development. The same cannot be said for national security-type problems. National security games are required to operate in an area where the questions and phenomena under study are in a constant state of flux – and by necessity this requires a different approach to game design and development. We believe that despite the complexity of the problem there are valid lessons from the simulationist approach which can and should be applied in a more qualitative context.

However, the complexities outlined in the preceding paragraphs do not absolve national security game designers from the responsibility of undertaking a more methodical and analytical approach to game design and development. Current practice involves an undocumented process of eliciting good ideas whilst using SMEs in the area of interest to advise the game team of underlying dynamics. This is a reasonable approach to take, but it should be accompanied by a literature review to establish a direct link between the phenomena that will be under examination and established practical or theoretic academic understanding.

The requirement for game designers to understand the phenomena they are designing the game to represent leads to a complex balance during the design process between knowledge required to undertake the task and the effort and expense of acquiring said knowledge. Without sufficient understanding the designers will not be able to effectively abstract the problem into game mechanics, and will therefore be unable to design an appropriate game system. However, game designers are not SMEs and the amount of time, effort and cost required to fully educate them in the many areas of interest to national security games would be prohibitive. An effective literature review in conjunction with SME support is intended to allow the designers to achieve a baseline understanding so that they can design an effective game construct when supported by SMEs who can bring depth of understanding.

However, at present the many assumptions made by both the SMEs and the design team regarding the dynamics under study and how they are being abstracted into the game's model and mechanics are not properly documented. Whilst the designers do produce documentation of the game's final ruleset – which often takes the form of detailed facilitation plans to direct discussions to area of interest – there is a clearly identifiable lack of transparency in the game's design that leads to an inability to explicitly state the assumptions that were made during its creation. If we are going to undertake literature reviews and properly employ SMEs it is critical that alongside this all aspects of the game's design are fully documented to demonstrate how this knowledge has been applied to solve the problem at hand. Proper documentation will allow the choices made by the designers in conjunction with the SMEs to be interrogated during post-game analysis; the underlying rationales behind game mechanics will be auditable and subject to scrutiny to prove that they were analytically sound.

We consider it entirely reasonable to accept that we will never be able to explain the underlying model of a discussion-based game to the level of detail which would equal a computerised model, but we can and should follow a more procedural and better documented approach to our design processes that is in line with computer modelling. This would take the form of undertaking further research into the phenomena in question, which should allow us to more explicitly link the game's design and mechanical choices to the evidence base. By extension, this would

increase our ability to explain why certain abstractions were made and to communicate their expected strengths and weaknesses more effectively to the customer. Doing this will provide more rigour in our documentation of the design process and require us to explain and justify the choices made by the design team during design and development.

In line with this, we outlined that it is good practice for an analytical game to have a Data Capture and Management Plan which should document the analytical framework of the game and the broader analytical process it supports. Due to the lack of planning and clarity regarding specific analytical methods applied to our previous national security games, in conjunction with the issues associated with capturing high-quality data as well as other constraints elucidated upon throughout this paper, we have generally not produced such a document when undertaking our national security games. This is an oversight that should be rectified in future efforts, especially if some of the recommendations made in this paper regarding data capture and analysis are acted upon. The DCMP is a critical document in ensuring a coherent and transparently justifiable analytical enterprise is maintained from the inception of the game design process to the delivery of the game's outputs to the customer.

Where possible the DCMP should state the current understanding of 'dependent' and 'independent variables' that are under investigation. A series of independent variables should be determined to represent the inputs or causes to stimulate players. The "dependent" variables represent the output or effects as determined by player decisions; the "independent" variables facilitate certain control variables that assist with the stimulation of the players as they work through their respective decision making and coordination activities. In our previous efforts we did not explicitly identify the dependent and independent variables under examination, and we believe this links to our observations that participants have sometimes erroneously drawn conclusions regarding variables that are independent or otherwise under the control of the game design team.

Where possible these variables should be established as early as possible in the game design process to inform the design, as game mechanics can/should be explicitly designed to capture information relating to the dependent variables to provide data for post-game analysis.[65]

Finally, when we design and develop games we need to keep in mind that analytical game design requires a conscious awareness that choices need to be driven by objective rigour, not just what 'works' in a game. Rather than putting mechanics into the design because they are immersive, interesting, or the easiest way to solve a game design dilemma, every element in the design should be present because we think it is the most appropriate way to illustrate the issue under consideration for analytical purposes – even if sometimes that might make the game design or delivery more complex or superficially less enjoyable to the players. This is more easily said than done given the constraints our games with seniors operate under as described

[65] We do however recognise that the expression of independent and dependent variables conveys firstly, the notion that we can postulate some kind of causal model, and secondly, underlying knowledge about the behaviour of the system to enable us to postulate said causal model. It must be acknowledged that this may not be the case when examining certain complex and wicked problems. This underscores that discussion based models inevitably produce qualitative insights that are open to interpretation, they represent the best method we currently have of generating insights into such unbounded and wicked problem spaces.

in section 4.2. Whilst there is an analytical imperative to design games purely with an eye to answering the research question, there are also practical considerations which make this difficult in practice. Senior players need to be engaged with the game's design, and there is a balance that has to be struck that will require compromise on both sides. However, we would argue that in our previous national security games we have failed to achieve the appropriate balance, prizing player engagement and immersion over the analytical purity of the design, as we are frequently driven by an unstated requirement to ensure good feedback on the event itself which is perceived as connected to the level of immersion and engagement. In future efforts the implementation of the recommendations put forward in this section can be used to help to redress this balance, in order to ensure that player engagement and immersion are not prioritised over the analytical purity of the design (or vice versa, as engagement and immersion help to maintain the quality of the outputs players produce).

3.4.1.1 The Evidence Framework Approach (EFA)

This section is a summary of P. V. Pearce, *A Technical Overview of the Evidence Framework Approach: Practical Ways of Thinking About Evidence,* DSTL/TR102712. Rather than simply referencing Pearce, this section will provide a fairly detailed overview of the EFA for two reasons. Firstly, the EFA in its current form is a key method we should employ, and we recommend its use as a method for being more analytical with regards to our post-game analysis in national security games. Secondly, a good understanding of the EFA will be required to support our subsequent recommendation that will be put forward in section 3.5, that a novel way to employ this framework would be as an evaluation tool for the *game construct* during design and development (in addition to its current use for judging the key insights that emerge from the game). This will allow the EFA to act as an analytical wrapper around the entire game process from start to finish.

Dstl designed and developed the EFA as a method of assessing evidence. The EFA is about practical ways to think about evidence and improving analytical quality, and helps people become better systems thinkers by undertaking an 'analysis estimate' process. This process provides a means by which people can consider evidence and its characteristics, and engage in discourse about evidence utility. The EFA comprises three tables used for evidence assessment – the Evidence Profile Table (EPT), the Validation Profile Table (VPT), and the Confidence Assessment Table (CAT).

The EPT is designed for use in assessing or evaluating the required or achieved quality of a body of evidence. Five factors are considered to be 'generic evidence characteristics', and are used to structure a conversation on evidence and to assign a profile that can be used to judge the extent of the 'warrant'[66] associated with an assertion or hypothesis. The profile can be scored between one and four for each of the five factors, as shown below in Table 6: The EFA's Evidence Profile Table
. The resulting evidence profile is simply summed to arrive at a statement concerning evidence warrantability, shown at the bottom of Table 6: The EFA's Evidence Profile Table
The five factors are:

[66] The 'warrant' is a general statement expressing the degree of belief in the quality of evidence for the hypothesis or proposition, ranging from 'Weak' to 'Beyond Reasonable Doubt'.

- **Comprehensiveness:** Considers the extent of the issues that have or will be explored and that are relevant to the proposition being made. This provides an indicator of the breadth and depth of coverage and understanding attainable. It also considers the degree to which uncertainties and errors have been or will be assessed and the extent to which coverage of the problem has allowed or will allow the system, its behaviours and its outputs to be understood.

- **Relevance:** Considers evidence drawn from a range of potential sources, e.g. previous studies, literature, data and assumptions and considers their relevance for informing the findings regarding the current problem. It also considers the extent to which sources have drawn on multiple relevant perspectives, and the extent of the inferential gap between assumptions and findings.

- **Challenge:** Considers the extent to which sources have been challenged and peer-reviewed by both the study team and relevant independent SMEs prior to wider exposure of the findings to customers. These aspects help to determine the extent to which the sources can be relied upon and how much challenge has been given to the findings.

- **Quantity:** Considers the number and variety of sources for generating the evidence. That factor also includes an assessment of the methods employed to generate the evidence, taking into account the number, scale and variety of approaches that have been used to tackle the problem. Where quantity is less of an issue, e.g. it is obvious how to proceed and ‘best-practice’ is available, it considers the extent of the track record for ‘best-practice’ methods in producing evidence. It is not necessary to have a large quantity of sources to score highly if there is a track record of appropriate use.

- **Veracity:** Considers the evidence in relation to the wider evidential picture, e.g. in terms of trends, patterns and explanation from across all or the majority of methods, the extent to which these form a highly supportive and integrated view and the extent to which alternative accounts for the findings are explored. This factor also considers what can be said about cause and effect.

Evidence Profile Table

Comprehensiveness Depth, Breadth	Relevance Perspectives, Assumptions	Challenge Peer Review, Scrutiny	Quantity Variety, Track Record	Veracity Relatability, Saliency, Causality	Profile Level
Extensive coverage of key issues & uncertainties; all behaviour explainable; 'known knowns'	Extensive artefacts & multiple perspectives; changing majority of assumptions does not impact; very small inferential gap	Extensive challenge, from department, national, international perspectives; caveats & assumptions clear, no limitations on utility for purpose	Extensive methods; subjective & objective; multiple alternative lines of enquiry or best practice approach with extensive track record	Proposition highly relatable to evidence base, supportive & integrated view; all relevant alternative accounts for findings eliminated; can state factor(s) A cause outcome(s) B	1
Majority coverage of key issues & uncertainties; some key behaviour explainable; 'known unknowns'	Good artefacts & some perspectives; changing majority of assumptions has limited impact; small inferential gap	Good challenge, from wider department; caveats & assumptions clear, some limitations on utility for purpose	Good methods; subjective & objective; many alternative lines of enquiry or good practice approach with good track record	Proposition largely relatable to evidence base, supportive & integrated view; alternative accounts for findings largely eliminated; can state factor(s) A very likely to cause outcome(s) B	2
Some coverage of key issues & uncertainties; some relationships can be described; 'unknown unknowns'	Limited artefacts & perspectives; changing some assumptions significantly impacts; large inferential gap	Limited challenge, from other relevant projects; caveats & assumptions clear, large limitations on utility for purpose	Limited methods; subjective or objective; few alternative lines of enquiry or single practice approach with limited track record	Proposition somewhat relatable to evidence base, supportive & integrated; some alternative accounts for findings eliminated; can state factor(s) A may well cause outcome(s) B	3
Majority of key issues & uncertainties not covered; very difficult to explain or describe anything; 'unknowable unknowns'	Very limited artefacts & perspectives; changing majority of assumptions significantly impacts; very large inferential gap	Very limited challenge, from within project; caveats & assumptions not clear, significant limitations on utility for purpose	Very limited methods; subjective or objective; no alternative lines of enquiry or single practice approach with no track record	Proposition have little or no relation to evidence base, not supportive & integrated; few alternative accounts for findings eliminated; can state factor(s) A might cause outcome(s) B	4

For a given proposition consider each criteria in turn. Select a cell in each column that contains the statements that best describe the situation, noting that not all statements within a cell have to be relevant. Assign a score based on the Profile Level. Once complete add up the scores for each criteria. Compare the total score to the Warrant Scale to derive a Warrant statement expressing the degree of belief in the quality of evidence for the proposition.

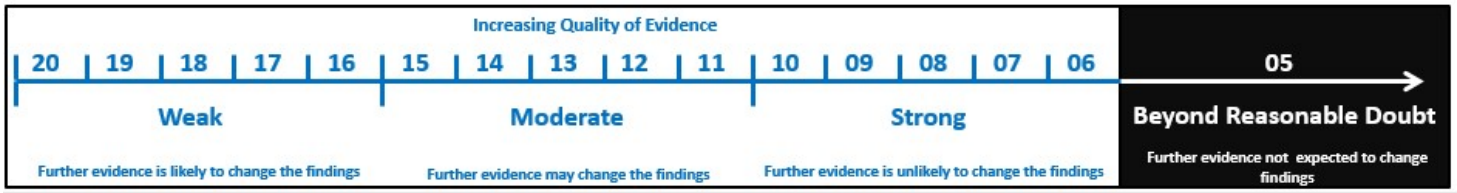


Table 6: The EFA's Evidence Profile Table

Using the above criteria the game team selects an appropriate profile level which they feel matches the quality of their evidence. Each of the statements within a level can be used to determine which cell in the EPT mostly characterises the assertion under consideration. It is not necessary to match every statement within a cell, and the cell which mostly characterises the issues considered should be selected.

The VPT is designed to complement the EPT. Its purpose is to assess or evaluate the validity of a body of evidence in relation to the proposition. The VPT is based on guidance contained within the Aqua Book.[67] The VPT allows a judgement to be made regarding the extent to which the right work is being or has been engaged in, taking into account the purpose and constraints placed upon that work. The key output from

[67] HM Treasury, *The Aqua Book: Guidance On Producing Quality Analysis For Government*, March 2015, UK OFFICIAL

this validation process is a judgement concerning the extent to which the work is valid as part of the 'fitness-for-purpose' judgement.

The VPT factors are used to structure a conversation on evidence validity and to assign a profile that can be used to judge the extent of the validity. They help understand the strengths and limitations of the analytical approaches. Specifically, the factors are:

- **Face:** Considers the extent to which the artefacts and supporting arguments for the proposition are considered relevant and plausible. On the face of it, do the outputs pass the "do I believe it?" test for the recipient?

- **Criterion:** Considers how well the evidence relates to the proposition being tested, and the extent to which the work actually engages with the issues that it claims to. This is about considering the extent to which the analysis has engaged directly with the relevant variables of interest or if it has used appropriate surrogates.

- **Construct:** Considers the adequacy of the game's mechanics in representing the issues under examination. This includes the key factors to which they respond and the mechanisms by which they do this.

- **Content:** Considers the extent to which it is possible to bridge the gap from data collected to genuine insights, as a result of its breadth, depth and granularity of evidence collected. This is also about considering if the analysis has measured and assessed the relevant aspects at the required level of granularity.

Each of the statements within a level can be used to determine which cell mostly characterises the validity of the assertion under consideration.

Validation Profile Table

Face Plausibility of artefacts	Criterion Appropriate measurement derived from artefacts	Construct Relevance of representation of key mechanisms	Content Interpretative Weight	Profile Level
Highly plausible; highly relevant & familiar to recipients; highly appropriate for intended purpose; highly relatable to prior experience	Key artefacts highly suitable for intended purpose; actual artefacts used for measurement; strong alignment between things being studied & thing being proposed	Key mechanisms highly appropriate; highly adequate & sufficient for purpose; strongly aligned to current understanding of issue	High interpretive weight giving extensive insight. Comes from strong focus on relevant issues & drivers with high fidelity in findings across breadth & depth	1
Largely plausible; largely relevant & familiar to recipients; largely appropriate for intended purpose; largely relatable to prior experience	Key artefacts largely suitable for intended purpose; surrogate artefacts used for measurement, largely adequate; good alignment between things being studied & thing being proposed	Key mechanisms largely appropriate; largely adequate & sufficient for purpose; largely aligned to current understanding of issue	Good interpretive weight giving good insight. Comes from good focus on relevant issues & drivers with good fidelity in findings across breadth & depth	2
Somewhat plausible, limitations with arguments; somewhat relevant & familiar to recipients; somewhat appropriate for intended purpose; somewhat relatable to prior experience	Key artefacts somewhat suitable for intended purpose; surrogate artefacts used for measurement, somewhat adequate; limited alignment between things studied & thing being proposed	Key mechanisms somewhat appropriate, some limitations with structure; somewhat adequate & sufficient for purpose; somewhat aligned to current understanding of issue	Some interpretive weight giving limited insight. Comes from limited focus on relevant issues & drivers with limited fidelity in findings across breadth & depth	3
Largely implausible; largely irrelevant & unfamiliar to recipients; largely inappropriate for intended purpose; largely unrelatable to prior experience	Key artefacts largely unsuitable for intended purpose; surrogate artefacts used for measurement, largely inadequate; no alignment between things being measured & thing being proposed	Key mechanisms largely inappropriate, major limitations with structure; largely inadequate & insufficient for purpose; largely lack alignment to current understanding of issue	Little or no interpretive weight giving little insight. Comes from little focus on relevant issues & drivers with little fidelity in findings across breadth & depth	4

For a given proposition consider each criteria in turn. Select a cell in each column that contains the statements that best describe the situation, noting that not all statements within a cell have to be relevant. Assign a score based on the Profile Level. Once complete add the scores for each criteria. Compare the total score to the Validity Scale to derive a Validity statement expressing the degree of belief in the validity of the proposition being made.

Table 7: The EFA's Validation Profile Table

Whilst the EPT assessment will result in an evidence score and the VPT assessment a validity score, there is often a need to express this in more simplistic terms to provide an overall assessment of confidence in the findings. This is achieved by using the CAT to cross-reference the warrant inferred from the evidence score and the validity inferred from the validity score. Both are used to make an overarching qualitative judgement about the confidence according to likely confidence bands. The confidence scale is “Very Low, Low, Medium, High and Very High”. Note that the confidence scale on the right of the CAT is shaded and not bounded by boxes – this is to illustrate that the boundaries are inherently fuzzy.

The summed EPT and VPT assessment scores are used to position the findings along the warrantability axis and the validity axis of the CAT. There is also a general

rule of thumb beneath each of the warrant criteria and adjacent to each of the validity criteria which provides a more informative statement about any judgement drawn in relation to these findings.

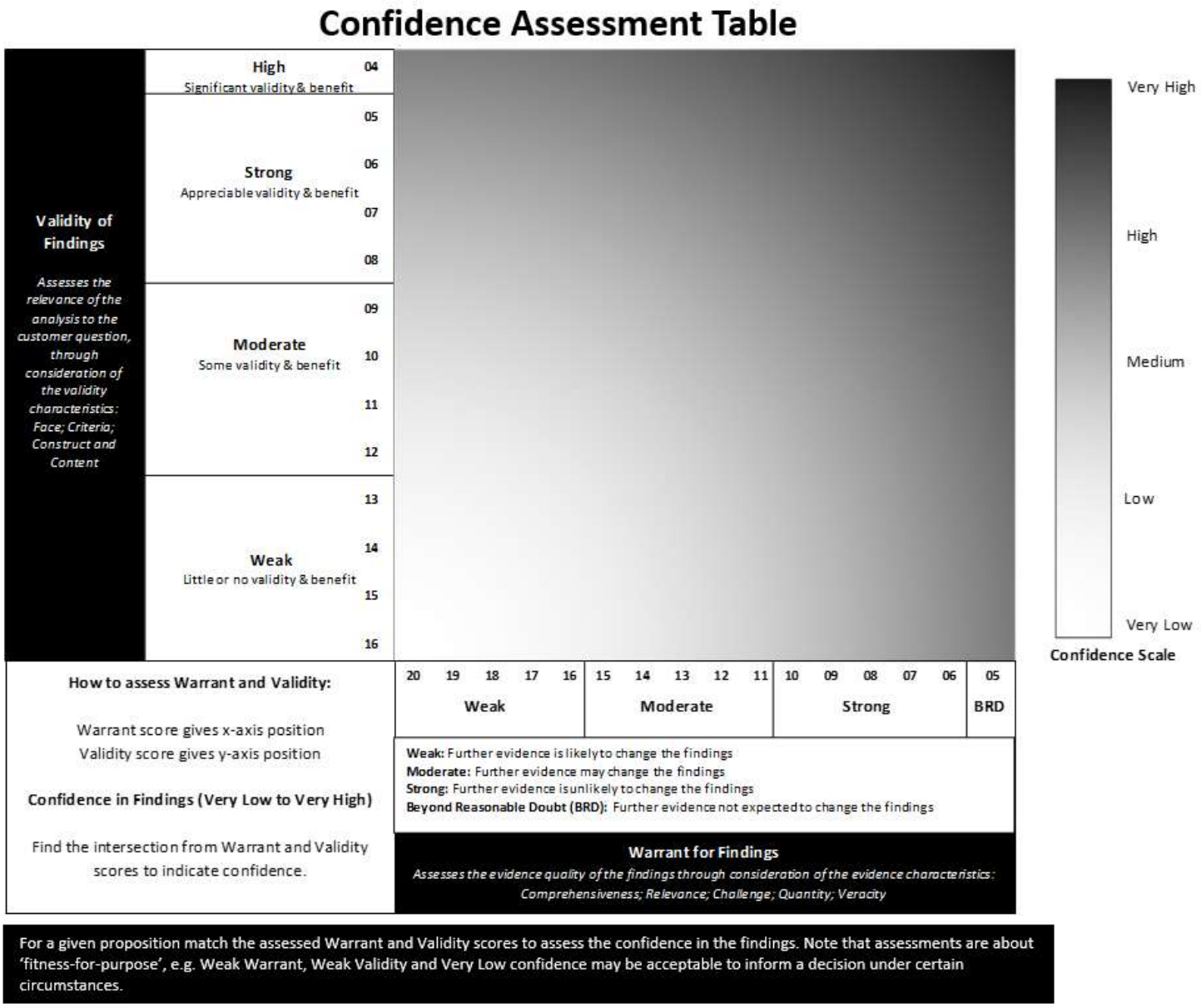


Table 8: The EFA's Confidence Assessment Table

The purpose of the CAT is to take the resulting profile scores and express the target or achieved levels of confidence. It should be noted that validity and confidence (the y axis) is a qualitative social construction and is likely to be fluid, i.e. there will be a number of perspectives, views, issues, etc. that need to be considered when seeking to make or understand a judgement about the agreement on the findings. Warrant on the other hand (the x axis) is likely to be fairly stable.

In summary, the EFA requires us as analysts to critique and evaluate the evidence generated to provide an assessment of our confidence in the findings that can be

communicated to the customer. It provides a method to clarify the strengths and weaknesses of the game's approach in a structured way without undermining the findings. Whilst it is very much a qualitative tool that produces subjective results it provides game designers with a method of assessing confidence in their results, which is something that is otherwise absent. The EFA is a significant improvement over the view that games are an art which cannot be measured or assessed. We would suggest that any game that purports to be analytical in nature would benefit from employing the EFA, and we would recommend that all future national security games should do so.

3.5 Verifying and Validating the Game Construct

In section 3.2 we identified certain aspects of the wargaming cycle that could be considered art and science. The ability of the game design team to abstract the game's underlying inputs into credible mechanics to create a functional game was identified as an area that is generally considered to be art, as it takes a game designer of considerable experience to accomplish this task. Sections 3.3 and 3.4 put forward a number of recommendations as to how the design and development process could be more scientific and analytical. This section will take this further, arguing that to be truly analytical during the design and development of a game the designers need to have a method by which they can analyse the game's actual construct to help verify and validate it. We believe this is necessary in order to hold games to a standard. Considering games as a pure art form also serves to effectively absolve practitioners of the need to consider qualitative analysis to the same standard as quantitative, and also neglects the genuine utility we can derive from effective national security games.

We contend that currently there are no analytical processes that we apply to analyse the game design, in contrast to the analysis applied to the outputs post-game. We believe that analysing the construct throughout the design process would allow designers to objectively assess whether their designs are fit for purpose prior to the game's execution. At present, such assessments are a product of simple expert judgement on the part of the game design team, and it could be argued that they lack structured underlying scientific procedures which would ensure objectivity. Identifying and applying such procedures would ensure that analysis is present throughout the entire cycle, and this follows best practice as shown in the wargaming lifecycle, Figure 3, which showed an analytical thread being present throughout.

Currently, we do not have any methods for undertaking such assessments in relation to our manual games – including all of our national security games. This section will argue that this gap can be filled by employment of the EFA, and will then provide a detailed overview of how this can be achieved. We would like to note at this point that we have consulted with Paul Pearce, the author of the EFA, as part of our research on this topic, and he has endorsed our proposals as a legitimate usage of the framework.

Our contention is that with minimal reinterpretation the criteria put forward in both the EPT and the VPT are just as applicable to the construct and the inputs that underlie a game as they are to the outputs, and that as the design and development process is

in progress the criteria put forward in both these tables can be used to assess the game's design as it develops.

At the start of the game design process the team should perform an estimate on the level of evidence required from the game. They should generate a series of EFA outputs which they would like the game to achieve with justifications as to why certain outputs are to be expected and should be achieved. Then, as the game design process progresses, the team should schedule slots to perform evaluations based on the criteria that are outlined in this section and compare them to the initial estimates. The results of these evaluations could either inform the design as it goes forward, and if changes are deemed necessary they will form part of the audit trail which explains why certain decisions were made. We will now show how the categories should be used by explaining how each of the individual criterion put forward in the two tables could be applied in this manner. This will be followed by a number of conclusions regarding the pros and cons of undertaking this approach.

3.5.1.1 The Evidence Profile Table Criteria

Comprehensiveness: *Considers the extent of the issues that have or will be explored and that are relevant to the proposition being made. This provides an indicator of the breadth and depth of coverage and understanding attainable. It also considers the degree to which uncertainties and errors have been or will be assessed and the extent to which coverage of the problem has allowed or will allow the system, its behaviours and its outputs to be understood.*

If adequate pre-game research is undertaken comprehensiveness can be employed during the design process to explicitly link this literature to an assessment of which areas of the problem space the model will (and will not) cover. This will force the designers to be explicit about the strengths and weaknesses of said coverage, and could potentially lead to identifying areas of improvement to ensure that the model will inform the question.

Relevance: *Considers evidence drawn from a range of potential sources, e.g. previous studies, literature, data and assumptions and considers their relevance for informing the findings regarding the current problem. It also considers the extent to which sources have drawn on multiple relevant perspectives, and the extent of the inferential gap between assumptions and findings.*

Relevance has two uses; firstly, it provides a check on pre-game research and inputs to ensure that the designers have undertaken this task properly and gathered a full range of relevant evidence upon which to base the generation of an appropriately realistic gamic model. Secondly it makes explicit and provides a check on, and if necessary a challenge to, the assumptions made during the creation of the model and their relationship to the game's outputs. This would help to explicitly identify areas where things are baked into the game design.

Challenge: *Considers the extent to which sources have been challenged and peer-reviewed by both the study team and relevant independent SMEs prior to wider exposure of the findings to customers. These aspects help to determine the extent to which the sources can be relied upon and how much challenge has been given to the*

findings.

When applied to the game design this would refer to seeking objective challenge regarding the model and its assumptions from other game designers outside the design team for the project. The would provide an objective third party assessment of whether the model is based on adequate research and if the abstractions made during the game design process have led to a product which is fit for purpose.[68]

Quantity: *Considers the number and variety of sources for generating the evidence. That factor also includes an assessment of the methods employed to generate the evidence, taking into account the number, scale and variety of approaches that have been used to tackle the problem. Where quantity is less of an issue, e.g. it is obvious how to proceed and 'best-practice' is available, it considers the extent of the track record for 'best-practice' methods in producing evidence. It is not necessary to have a large quantity of sources to score highly if there is a track record of appropriate use.*

Quantity it is partly interlinked with relevance. Relevance provides an assessment of the quality of your pre-game research, and this can be used to assess its quantity, i.e. are the model and its assumptions based on a comprehensive and thorough review of all the relevant literature and other materials that should have been engaged with to inform the game's design. Quantity mandates an objective assessment of the game's data capture methods – whether they are well-considered and substantive enough to provide the amount of evidence needed to execute the post-game analytical processes which have been identified in the DCMP.[69] Quality also considers the means by which certain game mechanics have been conceived and the data behind them created – have numbers or rules been created by SMEs estimating values, modifiers etc., or have a number of techniques been used to elicit such data, e.g. structured interviews, off line models, etc.?

Veracity: *Considers the evidence in relation to the wider evidential picture, e.g. in terms of trends, patterns and explanation from across all or the majority of methods, the extent to which these form a highly supportive and integrated view and the extent to which alternative accounts for the findings are explored. This factor also considers what can be said about cause and effect.*

When applied to the model, veracity asks whether there are competing theories that have emerged from research which could explain the phenomena under study and lead to conflict within the mechanisms present in the model, and if contradictions are identified it should lead to the designers making explicit how this has been resolved – has the game team chosen a specific theory to model, or have they tried to generate a model which would functionally satisfy a range of competing theories? It also asks

[68] It is important to note that at this stage of the process the game has not yet been play tested (which is where the VPT comes in) so this is about challenging initial game design decision and how the games will hang together prior to play testing.

[69] The idea of quantity at this stage of the design process is to look at it from an analytical perspective; considerations would be about the 'quantity' of approaches to capture observations, insights and data during game execution with multiple methods for capturing multiple perspectives better. Alternatively, there may be a single best practice approach for capturing data during certain types of game.

whether factors that have not been accounted for in the game's design could be significant explanators, and the game design team will use this to argue/explain exactly how the model can be used to draw valid conclusions that are relevant to the phenomena under study.

Veracity should also allow the designers to consider the game design in terms of the outcomes that could happen; i.e. if the dependent variable(s) are about the effects observed, does the game design support assessments of indirect and direct evidence? Indirect and direct evidence are defined as follows:

Indirect: Evidence that helps eliminate alternative causes for the effect observed.

- If an outcome occurred in the scenario, does the game support exploring the issues around the outcome to determine if it is possible to rule out anything other than the intervention – such as changes in the starting conditions – as the cause of the outcome in the scenario? If so, then the game can support the premise that the intervention must have caused it.

Direct: Evidence that specifically relates the cause to the effect observed. Consider:

- **The character of the effect** – Does the game design allow analysts to track and trace an intervention to see the effect? In essence can the game support determining if the effect occurred at the time, in the manner and of the size to be expected if the intervention had caused it?
- **Symptoms of causation** – Does the game design allow for tracing and tracking the side effects that could be expected had the intervention operated to produce the outcome? Does the game allow analyst and players to see ripple effects that could only be explained by the intervention?
- **Presence of requisite support factors** – Does the game design support understanding the state at certain time to determine if everything was in place that needed to be in place in order for the intervention to produce the outcome?
- **Presence of expectable intermediate steps** – Does the game design support determination of the sequence of events that preceded the effect being observed to determine if the right kinds of intermediate stages were present?

3.5.1.2 The Validation Profile Table Criteria

Face: *Considers the extent to which the artefacts and supporting arguments for the proposition are considered relevant and plausible. On the face of it, do the outputs pass the "do I believe it?" test for the recipient?*

Face can be used as a check on whether the model is providing outputs that seem both plausible and reasonable in terms of explaining the dynamics under study. This evaluation would be applied after playtesting has taken place to make sure that the design team properly evaluate the outputs from said playtest(s). Face helps to ensure that the game will function as expected.

Criterion: *Considers how well the evidence relates to the proposition being tested, and the extent to which the work actually engages with the issues that it claims to. This is about considering the extent to which the analysis has engaged directly with*

the relevant variables of interest or if it has used appropriate surrogates.

Criterion confirms that if the customer wishes the game to make a judgement regarding X then the designers will need to explain the mechanisms that either represent X appropriately or can be employed as an analytically justifiable and defensible proxy/surrogate for X. This explicitly identifies the link between the game's mechanisms and the variables under study in a way that we currently do not.

Construct: *Considers the adequacy of the game's mechanics in representing the issues under examination. This includes the key factors to which they respond and the mechanisms by which they do this.*

In this context construct focuses specifically on the nuts and bolts of the game's design, at the heart of what most would call the art of gaming. As with 'criterion' above, evaluating the construct also focuses specifically on the mechanics of the model. However, in this case it would be to assess the appropriateness of the selected gamic mechanisms, building on the previous assessments of both 'relevance' and 'criterion'. These focused on how the research was condensed and abstracted down to inform the mechanics of the game design and examine the correct variables. A construct evaluation ensures that once the mechanics have been designed and implemented they are appropriate for their purpose and still maintain a direct link to the inputs.

This means the design team have a fully auditable trail that makes certain the game actually engages with the issues that it was designed to and directly links the mechanics to the variables under study.

Content: *Considers the extent to which it is possible to bridge the gap from data collected to genuine insights, as a result of its breadth, depth and granularity of evidence collected. This is also about considering if the analysis has measured and assessed the relevant aspects at the required level of granularity.*

Whilst 'face' assessed whether the likely outputs were plausible in nature, content will inform the designers as to whether the game is likely to be able to gather the necessary data – both in breadth and depth – to facilitate the execution of the analytical methods proposed in the DCMP. This will allow a final assessment of whether the game's outputs will be adequate to actually provide the customer with a *better informed* answer to their question. In conjunction with this, content should provide the first in a constant series of checks to ensure that the game's output will a) deliver valid insights, and b) that these insights are likely to deliver real benefit to the customer.

We think that evaluation sessions undertaken during the game design process using this modified framework would provide three main benefits:

1) A structured framework to evaluate game designs
Until this point we have not applied any scientific methods to evaluate the game model and the design process, and we believe that using the EFA for this purpose will ensure that our national security games will be more analytical. This assessment would be embedded into the game design and development process, with different criterion being relevant at different points in time. We do not see this evaluation as needing to take place in a single session, rather, the individual criterion should be

applied as and when they are relevant in the process. The outputs of this evaluation could then be used to inform changes in the game's design to address areas of concern, thus increasing the quality of the end product.

2) Transparent and Auditable Game Designs

Employing this structured framework will lead to transparent and auditable game designs. Assumptions will be made explicit and throughout the process decisions taken will be linked directly back to the underlying evidence informing the model. In conjunction with proper documentation and a design diary this will form part of an audit trail allowing a post-game assessment of what decisions were taken and why – leading to much better understanding of the strengths and weaknesses of the model and the analysis, as well as an improved method of communicating this to the customer.

3) Maintaining standards of good practice

An additional benefit to employing this approach is that it will ensure the maintenance of good gaming practice during game design and development. If the game team are unable to fully address the criteria required by this framework this may suggest that they have either skipped or paid insufficient attention to aspects of the game design process, which could lead to problems during execution or analysis.

The down-sides of implementing this approach must also be acknowledged. As with all analytical processes it would take substantive effort to implement, and this would translate to increased development times and costs. This would be a laborious increase on the work required by the game design team, as they would need to capture data on and document considerably more of their processes than is currently required. Whilst the analytical value of this exercise may be clear to the analyst, it may be difficult to justify increased development time and expense to the sponsor, especially given that from a player perspective it has no obvious perceptible impact on the game's execution which is often where their area of focus lies. Implementing this procedure requires that the customer is educated in and understands the benefit it will bring so that their buy-in can be assured. However, it is also important to note that it is part of our ethical duty of care as MOD scientists to provide quality assurance, and so customers should be willing to accept some additional cost as part of good practice in gaming. Furthermore, if Dstl's Defence Wargaming Centre is aiming to become a recognised centre of excellence for gaming then we will need to move forwards and break the boundaries we have identified on innovative analytical practices.

3.6 Data Capture and The Employment Of Analytical Methods For Post-Game Analysis

One of the areas of improvement identified in section 3.2 was the DCMP. This section will focus on part of this plan – the collection of data. A game needs to capture appropriate data that will be utilised as part of a considered analytical process to produce meaningful and valid insights that are relevant to the stakeholder's stated question or requirement. What data is appropriate will vary within the context of each specific game, but it will always consist of data *that is of sufficient quantity and of the right type* to allow the analytical methods/techniques that were pre-selected during the game's design to be properly executed.

There are two main routes through which better access to appropriate data could be exploited.

3.6.1 Exploitation Routes

The first exploitation route is comparatively simple – better access to appropriate observational data could be used to employ our current methods to greater effect. Currently the observational data we collect from our national security games allows us to reliably construct 'what happened', i.e. the actions players decided to take in the game, but it is not as useful in explaining 'why it happened'; we do not collect enough relevant data from participants pertaining to their rationales to provide validated insights into the decision making processes that led to the observed actions. Access to more appropriate data in both these areas would enable us to compare and contrast a range of different views from the participants both as a group and as individuals, and this would provide a measure of challenge to individual viewpoints. The authors contend that these comparisons will allow us to consider our insights to be more valid, and that in combination with information as to why events happened they will also be inherently more meaningful.

The second exploitation route is more complex and nuanced, and will be the primary focus of this research. Better access to appropriate data is an enabler of post-game analysis through a range of potential analytical methods. The authors then assert that appropriate analysis of the data could lead to more meaningful and valid insights being produced, in comparison to reporting observations. Furthermore, these insights will be produced by members of the game team – who are fully informed of the game's events across all the cells, as well as its mechanical strengths/weaknesses – rather than being based on observations by participants who have a valuable but more limited individual perspective.

3.6.2 Capturing Data – Increasing Quantity and Accessing Different Types Of Data

3.6.2.1 Current Data Capture

Post-game insights from previous national security games have been primarily generated from a reconstruction of the game's narrative in conjunction with observations from the game team and the SMEs who were present – the data captured has thus far not been subjected to rigorous challenge or specific analytical methods as part of a well-defined analytical process.

When designing a qualitative Discovery Game there is a complex interrelationship between:

a. the proposed analytical method informing the quantity and type of data that needs to be captured; and,

b. an assessment of the quantity and type of data the designer thinks it is realistically possible to capture informing which analytical methods will be viable.

In previous national security games data we have primarily taken the latter approach, with data capture being implemented without a full consideration of analytical methods during the game's design process. This has led to a somewhat post-hoc

approach to analysis based on capturing data and subsequently deciding what – if any – forms of analysis can be applied to it.

There are two overarching types of data that can be captured during a game. The first is game state data, which provides a snapshot of the actions taking place within a game's mechanisms at a specific point in time.[70] The game team needs to capture the game state accurately to run the game and generate credible outcomes – and by extension a credible game narrative. A facilitator will lead each player cell through a schema/framework that the participants follow to capture the actions taken each turn. This data needs to be captured in a codified/structured format that is suitable for processing by the game's adjudication processes, and this allows an outcome to be generated for the turn in question. The capturing of this game state data also allows for a detailed post-game understanding of the events that took place in gameplay, and this is something that currently our national security games successfully capture to a comparatively high standard.

The second type of data captured is player domain data, which encompasses any data which can be mined from players and player discussions in cells during or after the game. Capturing player domain data properly is a complex task in national security games given that they are freeform environments where many discrete conversations can be taking place concurrently. Thus far a small number of data capturers have been tasked with taking general notes pertaining to conversations between players. This task has not been structured and is often under-resourced given the complexity of the environment. Data captured is of a comparatively low standard, and a significant amount of potential player domain data is not captured. The loss of the data is key in explaining why we have struggled to reconstruct *why* actions were undertaken and to understand player rationales. When data is collected there is also an overreliance on post hoc assessment – after the game, players have been asked to explain the reasoning behind key decisions during plenary sessions, but this takes place after they have seen the outcomes and is likely to result in some degree of bias entering the responses.

Our current view is that we are relatively content with the quantity and quality of the game state data that is captured, and we believe that we can recreate an accurate picture of the events which took place in gameplay and by extension construct an accurate narrative of events post-game. Subsequent discussion will therefore primarily focus on player domain data, as this is an area we have identified where we are currently missing data and different types of data are available which we have yet to properly explore.

3.6.2.2 Caveats

Before we discuss specific methods of increasing the quantity of data and accessing different types of data we would like to state two caveats.

Firstly, for the purposes of this paper we will discuss a range of best practice theoretical options which could fix these problems, initially disregarding practical

[70] The game state changes whenever any of the objects or values in the game are modified, moved or manipulated by the players. Such changes can denote either an action that was taken, or potential courses of action that were under consideration at the time but which were ultimately rejected. Game state data is generally quantifiable and can be employed during post-game analysis to reconstruct a narrative of events that took place.

constraints. We will consider the implications of such constraints later in this paper, as it will be obvious to informed readers that some of the suggestions will be practically difficult to implement. However, we believe a discussion of best practice options will be of value, as only once we understand the ideal situation can compromises that provide 'best effort' solutions be properly understood and formulated.

Secondly, the data capture methods suggested below and given as examples are by no means meant to provide a comprehensive list of all the possible methods we could potentially employ and their pros/cons. We have selected specific examples which best illustrate the points in question and provide food for thought for game designers/analysts.

3.6.2.3 Capture Methods to Increase Quantity and Access Different Types of Data

Data capture methods are generally considered to be either passive or active.

Passive data collection involves data gathered without directly engaging the participants, in a manner which is mostly or entirely non-intrusive from the perspective of the players – examples include ethnographers taking notes of conversations or audio recording. In gaming terms passive data collection does not impact on the game's mechanics and requires little to no change in how the game would be executed from the perspective of the players.

We have already employed passive methods to capture player environment data in our national security games via the use of ethnographers. However, even in large player cells we have normally tasked a single analyst or a small number of analysts to undertake this role. The reason for this is twofold. Firstly, data capture requirements are to some extent driven by the requirement to employ certain analytical methods. In most of our national security games up until this point there has not been a formalised requirement which would have driven the collection of substantive player environment data. This has been due to a lack of interest from our customers/stakeholders for game outputs that are supported by analysis; the value of the event has been seen by them to be experiential and observational, and stakeholders have considered player discussions in post-game plenaries as adequate to provide good enough data from which to generate insights. Part of the reason why we are conducting this research is that we disagree with this line of argument. Secondly, this data requires a concerted effort backed by a substantial resource investment in manpower and time to properly collect and analyse it so that meaningful robust insights can be drawn. Given the lack of a driving requirement for post-game analysis this would have been hard to justify to our stakeholders in previous games.

Without any change to our current methods we recognise that an additional resource and manpower investment in ethnographers would allow them to provide a substantially greater coverage of discussions during the game's execution. This would substantively increase the quantity of player domain data collected, and contribute towards capturing some data which has thus far been lost; the expected value of this would be that capturing player discussions would increase our understanding of the decision making processes which led to the actions that were taken.

Although this paper will not go into detail on this topic, we could also employ technological enablers to enhance our passive data capture, examples of which include microphones for audio recording, cameras for video recording and sociometric badges.[71] Technological enablers would allow for higher fidelity data capture and would provide access to different types of data – especially in the case of sociometric badges – but careful consideration would need to be made as to whether this data would usefully contribute to generating the kinds of insights we seek to draw from these games. This is why the proposed analytical method should inform the quantity and type of data that is captured; simply capturing more data is not in itself helpful, there must be a clear understanding of how it will be employed. Within the context of our national security games, considered use of audio recording could provide useful data and also help to alleviate the requirement for additional ethnographers to be present during the game's execution.

One of the greatest advantages of appropriately trained ethnographers is that they can parse conversations as they take place, and this will allow them to collect only the data that is relevant to answering the questions at hand. A well-trained ethnographer will be fully appraised of the game's goals and analytical method as well as having an expert understanding of the topics and concepts being gamed. They will also be experienced in – and empowered to – undertake on-the-spot interviews during the game to interrogate players' actions and the rationales behind them. They need to be able to read the room to understand and recognise key data capture opportunities where important moments are taking place and insightful statement/views are being put forward. Such moments are where analysts post-game will need to know exactly who made certain comments, as well as the context. Currently, it is not routine practice to train our ethnographers – we assign this task late in the process and assume that analysts already possess the relevant skills and background knowledge. This assumption is often faulty and leads to highly variable quality of post-game notes depending on the experience of the individual analyst in question. The problem is further exacerbated by a tendency to assign this task to junior members of staff as our more experienced gamers are generally required to run or take part in the game. The authors recommend that formalised training for ethnographers should be undertaken to increase the quality of data captured.

We can also consider implementing additional methods/mechanics in the game that could be employed to generate different forms of data which are amenable to meaningful analysis and can be collected passively. As an example, the integration of ordinal scale trackers. In a game seeking to examine escalation (which has been a topic of national security games that we have run, and continues to be a topic of interest) when an action is taken by the players, members of the adjudication cell

[71] Sociometric badges are a hardware based solution involving lanyards worn by players. The badges can capture data relating to face-to-face interaction, conversational content and dynamics (e.g. turn taking patterns, tone of voice, etc.), physical proximity to other people, and physical activity levels using social signals derived from vocal features, body motion, and relative location.

could be tasked with evaluating the action on an ordinal scale – such as a Likert scale[72] – a hypothetical example of which is shown below:

Do you consider the action in question to be:

- *Very Escalatory*
- *Somewhat Escalatory*
- *Not Escalatory or De-escalatory*
- *Somewhat De-escalatory*
- *Very De-escalatory*

This would potentially allow us to access new and useful data, as currently we have not successfully drawn inferences regarding the comparative level of escalation of different actions taken throughout a game. A well-established range of social scientific methods/techniques can be applied to analyse Likert scales, and as part of our inductive method such analysis could contribute to the generation of further hypotheses surrounding escalation dynamics.

Active data collection of player domain data is gathered directly from the participants either as groups or individuals through some form of engagement; examples include interviews and surveys of participants. Active data collection methods need to be supported and facilitated by the game's design, and the game will therefore to some extent need to be built around them. These methods may change gameplay or how the game is executed from the perspective of the players – for example, the game may need to be paused so data can be collected at specific times or when critical decisions need to be made.

Active data collection of player domain data has been employed only minimally in our national security games thus far and has been limited to semi-structured post-game plenary sessions. Whilst the data collection methods used in these sessions have primarily still been passive note taking of discussions, the overall plenaries can be considered a form of active data capture as they take place outside of gameplay, are led by a facilitator, and are built around a structured series of pre-determined questions the game design team would like the participants to discuss.

Active data collection will allow us to access types of data we have not engaged with so far. However, the consequences, costs and benefits will need to be considered, as these methods will generally take additional time and require gameplay to be paused – this can break player immersion and damage the flow of the game. The game designers therefore need to weigh the pros and cons of employing such methods carefully during the game design process.

Focussing at this juncture on the pros/advantages, embracing active data capture methods will allow us to interview players at specific points throughout the game's execution. If ethnographers were empowered to pause the game to ask probing questions to either groups or individuals this could provide much more detailed exposure of the rationales, assumptions, and decision making processes of the

[72] A Likert scale assumes that the strength/intensity of an attitude is linear, i.e. on a continuum from strongly agree to strongly disagree, and makes the assumption that attitudes can be measured.

McLeod, "*Likert Scale Definition, Examples and Analysis*", Simply Psychology, 2019 https://www.simplypsychology.org/likert-scale.html

players. This would enable a much more detailed post-game assessment and analysis of the decisions made and more importantly the influences/drivers of those decisions.

Ordinal Likert scales could also be employed to great effect as part of an active data collection plan. As and when decisions are made, players could fill out forms to capture their views on the action(s) which have been agreed. This would provide a significant stream of additional data which has thus far not been tapped – as elucidated above. Whilst forms would need to be pre-designed, we could also employ e-voting to ask further questions on-the-spot and address decisions/topics of interest as they take place (furthermore, e-voting is likely to be less disruptive to the flow of the game in comparison to filling out forms). This would provide data capturers with a method to engage players dynamically and anonymously, and provide immediate results that could be displayed to engender further discussion if required. In a group setting a requirement to fill out such scales could also be employed as a framing tool to facilitate discussions around topics of interest. If the group is required to agree on, for example, how escalatory their actions are this will spark debate on an area of direct relevance to the question at hand, building into the game's design an opportunity to collect relevant data.

As a note of warning, in reframing discussions around such paradigms players are potentially no longer engaging with the game entirely realistically, as their thinking about the situation at hand may be altered due to their interactions with an artificial conceptual framework. It is possible that the act of engaging in such discussions could in itself change the player's thought processes and, potentially, the actions they take, corrupting the results and leading to less realistic decisions. This effectively asks the conceptual question of the degree to which observing a game can change the output of said game. This research will not attempt to solve this problem, but will put it forward as a point to be considered by the game's design team when engaging with active data capture methods and designing their data capture plan. We recommend that the impact of active data capture methods on gameplay will need to be assessed during the game's playtests to either ensure that they have as minimal an affect as possible or to make explicit the areas where such an impact could occur so as to make this issue explicitly recognised during post-game analysis.[73] Additionally, some playtests should intend to produce a credible narrative to allow for data capture and analysis processes to also be tested – this will allow the design team to be confident that they will be able to generate the required output when writing the post-game report.

3.6.2.4 Practical Constraints

When generating the data capture plan for a game, the design team will need to consider a broad range of practical constraints that will affect data capture efforts, many of which have already been alluded to. During the execution of our national security games we have further identified which specific constraints are of significance for data capture and analysis.

[73] However, we must caveat that for various reasons including (but not limited to) lack of player availability leading to proxy players being used and the knowledge that this is not the actual event, playtests are a flawed means for generating outcomes that are likely to match those of the actual game's execution. Therefore, there are limits to what a playtest can tell us.

The first constraint is time. This includes both time required to collect data during the game and time required to analyse the data. As discussed in the previous section active data capture methods interrupt gameplay to engage the players for data capture purposes. However, games run with senior participants are often already compressed into extremely tight timeframes – the shortest thus far being four hours. This already presents a major obstacle to be overcome by the game's design team in terms of running a viable game. An additional requirement to spend some of this limited time on active data capture activities which disrupt gameplay would undoubtedly be challenging to implement.

Post-event the collection of additional data will also require further time to catalogue and analyse, and this can be particularly true in relation to qualitative data. One game run in the US which required participants to fill out surveys over three days resulted in 'a stack of completed surveys about 3 feet high. It took 2 junior staff members a month just to input it all into a database'.[74] Furthermore, the weakness of using technological enablers when employed to passively collect data is that they are generally indiscriminate, and therefore would need to be employed judiciously to avoid information overload. Simply attempting to record the entire game will not be a viable solution in most cases. Even our better resourced US counterparts state that such an approach was 'dropped soon after being tried due to the lack of time and appetite for listening to, transcribing and wading through 5 days × 8 hours play × number of cells of recordings. To say nothing of senior officers being too humble to want their wisdom captured in such an attributable way'.[75] However, in the context of senior games that are generally time compressed the practical viability of using audio recording may actually increase, as a short game will generate comparatively less data that requires analysis.

The second constraint is, as stated above, that senior audiences can be reticent to have their conversations recorded due to strong feelings regarding direct attribution of discussions, and this often applies during the context of games where participants are required to conjecture over worst case scenarios or to choose between multiple unpalatable options. This could potentially lead to participants being unwilling to engage with the game due to the unacceptable nature of the data capture solution. Difficulty gaining senior buy-in for the employment of audio recording is matched by the practical difficulty in obtaining audio recording equipment that is cleared for use in high classification environments. These factors have thus far combined to rule out the use of technological data capture solutions. We recommend that in future if a requirement is identified to use technological methods to capture data in a situation where they would bring an identified benefit then they should not be dismissed due to these issues. We should attempt to engender a cultural shift in our customers so that they begin to accept such methods as routine.

The third constraint is in the amount of resource required to collect, catalogue and analyse the data in terms of manpower. Comprehensive data capture greatly increases the number of ethnographers required to participate in the event. In one analytical US game 'There were approximately 40 players in the game, and maybe 15 note takers'.[76] Once collected this volume of data also required both time and manpower to catalogue and analyse. This creates a substantially increased burden to find and train enough personnel to undertake these tasks, and if personnel can be

[74] Compton, Email Correspondence with Author, 14 Apr 2020
[75] Downes-Martin, Email Correspondence with Author, 7 Apr 2020
[76] Compton, Email Correspondence with Author, 14 Apr 2020

found they will come at substantial cost to the project – some customers would likely balk at this ask. A cost/benefit assessment would need to be made by the design team to recommend to the customer a reasonable amount of data capture to fulfil their requirement, and it is inevitable that the methods used will ultimately be guided by what is feasible in terms of cost whilst being fit for purpose to address the issues. Once again, careful consideration will need to be given to resource and cost during the game's design so that the impact of decisions can be communicated to and understood by the customer.

3.6.3 Employment of Analytical Methods During Post-Game Analysis

3.6.3.1 Current Data Analysis and Reporting

Our four previous national security games have all followed a fairly similar trajectory in terms of how the data collected was analysed to draw insights that were reported to our stakeholders.

Firstly, we reported the narrative that was created. Our understanding of the narrative was based on the actions that the player cells took over the course of several game turns. These actions were recorded by the cell facilitators following a pre-determined framework, and some further understanding of the wider context was integrated via the notes of ethnographers who were present for the player cell discussions. Secondly, facilitated semi-structured plenary discussions during the event – generally after it is completed – were used to collect player observations into areas of relevance. Analysts sanity checked these observations using their own judgement, and the player observations then formed part of the evidence base for the conclusions from the event. Thirdly, we drew conclusions based on the narrative and player observations via the best judgement of our analysts; this process did not follow a codified, structured or formal method. Finally, these conclusions were presented in a formal report and were backed up by insightful quotes from players. Due to the high level of interest in these games this report was normally a flash report due for immediate delivery in the weeks after the event, and as such generally took the form of a short series of key observations/insights presented over a handful of sides of A4.

Our current view is that this method allows us to produce a set of insights into the customer's problem rapidly. However, our view that being more analytical in our approach will serve to generate more meaningful and valid insights stems at least in part from our concerns over the limitations of this process.

3.6.3.2 Enabling Different Analytical Methods

Throughout this section it has been asserted that the collection of different types of data would enable analytical approaches/methods that we have thus far been unable to employ. This subsection will outline these approaches/methods and examine the implications of their use.

Collecting this additional data could enable a number of qualitative analytic approaches for use where appropriate such as content analysis and grounded theory, as well as potentially discourse analysis and narrative analysis:

Content Analysis: Described as "a ...method whereby a researcher seeks objectively to describe the content of communication messages that people have previously produced." "Content analysis involves identifying coherent and important

examples ... and patterns in the data ... [and subdividing] ... data into coherent categories, patterns, and themes."[77] Content can be sorted into "bins" to determine which, if any, of the focus areas are supported by participant comments or ethnographer observations.

Grounded Theory: A more detailed, methodologically sound approach to analysis than the initial step of content analysis, grounded theory employs systematic, hierarchical procedures to develop inductively derived theory grounded in data. Grounded theory "directs researchers to look for patterns in data so that they can make general statements about the phenomena they examined."[78] It is possible to use both deduction and induction when utilising this approach, thus allowing the team to employ a "... theory discovery methodology that allows the researcher to develop a theoretical account of the general features of a topic while simultaneously grounding the account in empirical observations or data."[79]

Thematic Analysis: Another particularly appropriate method (or suite of methods) to apply is thematic analysis. Thematic analysis identifies 'patterns of meaning across a dataset that provide an answer to the research question being addressed. Patterns are identified through a rigorous process of data familiarisation, data coding, and theme development and revision'.[80]

The authors have employed this method of analysis outside the national security gaming space, to analyse the output of other qualitative discussion based games. When running the Royal Air Force's (RAF) *Homeland Resilience* wargame the player's discussions were recorded in full using digital audio recording, and these were then transcribed in full to enable a thematic analysis.[81] Thematic analysis can be undertaken as an inductive approach, and as such the process of coding occurs without trying to fit the data into pre-existing theories or frameworks. The volume of data collected allowed the analysts to identify patterns and to extrapolate various hypotheses from the bottom up based on a substantive body of evidence. Employing this method helped to make sense of a large body of data and to avoid an overreliance on a very small sample set of seemingly insightful observations from individuals during post-game plenaries. Thematic analysis also provides a method to reasonably evaluate conflicting observations raised by participants through a considered evaluation of weight and quality of supporting evidence. Critically, although player observations formed the majority of the data this approach allowed the analysts to draw substantiated conclusions from these observations, as opposed to simply stating observations that were judged to be insightful which could be the result of groupthink.

---

[77] Levin, Fox, "*Elementary Statistics in Social Research"* (New York: HarperCollins, 1991), 6; Patton, p.149

[78] Potter, *An Analysis of Thinking and Research About Qualitative Methods* (Mahwah, NJ: Erlbaum, 1996), p.151.

[79] Myers, "Qualitative Research in Information Systems" in Patricia Y. Martin, and Barry A. Turner, "Grounded Theory and Organizational Research," *The Journal of Applied Behavioral Science, (22:2)*, 1986, pp.141-157

[80] The University of Auckland, "Thematic Analysis: A Reflective Approach", https://www.psych.auckland.ac.nz/en/about/thematic-analysis.html

[81] During this process many of the constraints identified above were evidenced, as the transcribing and codifying of the data took entire weeks of an analyst's time to produce over 2,000 separate player observations.

Moving beyond a pure thematic analysis approach into grounded theory allows analysts to generate and state their own conclusions and insights drawn from the data – rather than simply reporting those of the participants. This provides a means to identify genuine insights and avoid the problem of 'insights' drawn by players that are actually a product of mechanics/dynamics that were baked into the game's design. As patterns and hypotheses are identified and generated any which are spurious will be transparently identifiable to the analysts working with the underlying data, as they have a full understanding of the game system's limitations. They will therefore have a methodologically sound justification – beyond simply not agreeing with a particular observation using their own judgement – to either exclude any such hypotheses/insights from the final report, or address them directly as to why such insights are not valid. This would also help to prevent/counteract negative learning from participants who have made their own judgements regarding the game's outputs from their experiences during the event. Furthermore, these methods will also allow the analysts to put forward insights that in some cases directly contradict the views put forward by individual participants. Individual players are often not in a position to see the complete picture of interrelationships between mechanics/cells that may provide a genuine explanation for why certain events have transpired. With access to more data that has been subject to analysis the game's analysts will have a sound and defensible platform from which to contradict player judgements.

Specifically in relation to analysing Likert-scale survey responses **Descriptive Statistics** can be employed utilising measures of central tendency such as the mean, median, mode, standard deviation, variance, and skew. These could be employed to provide some degree of quantifiable assessment relating to player views on topics that are generally considered to be highly qualitative – such as escalation. This would also allow for codified comparisons between players, actions within a single game and games between each other.

Finally, the last analytical methodology we would like to highlight is **Q methodology**. Q methodology is a research method used to study subjectivity. Normal factor analysis 'involves finding correlations between variables... Q factor analysis reduces the many individual viewpoints of the subjects down to a few "factors," which are claimed to represent shared ways of thinking… ranking, rather than asking subjects to rate their agreement with statements individually, is meant to capture the idea that people think about ideas in relation to other ideas, rather than in isolation'.[82] Q methodology can quantify the degree of consensus in a group, and it also allows for some analysis to take place during gameplay as player surveys can be codified live – which could potentially directly involve the participants in the analysis.

If data capture improvements as put forward in the previous section are realised then we should closely consider the employment of technological enablers during post-game analysis. Our colleagues in the US have previously used the ATLAS.ti software package to help with codifying and analysing data collected from qualitative games. ATLAS.ti can be used to help researchers uncover and systematically analyse complex phenomena hidden in unstructured data (text, multimedia, geospatial). The software provides tools that let the user locate, code, and annotate findings in primary data material, to weigh and evaluate their importance, and to visualize the often

[82] Q Methodology https://en.wikipedia.org/wiki/Q_methodology

complex relations between them.[83] It also provides 'analytical and visualization tools designed to open new interpretative views on the material' including visual model building and 'mind mapping', proximity analysis of coded data and integrated visualizations.[84] The software's utility is that it facilitates analysis by increasing the ease and effectiveness of coding large amounts of qualitative data over different forms of media. Engaging with audio/video data would now be viable – currently this cannot be properly handled by our current approach, which involves coding text within an Excel document, and transcribing data from these formats reduces its richness. The use of such software can help to increase the speed and efficiency of cataloguing, searching and making sense of the data, as well as undertaking potentially useful analysis on it. Coding can even potentially be automated, which could help to address some of the constraints related to volume of data when undertaking qualitative analysis. Ultimately, the use of such tools can actively contribute to the analytical methods we can employ, providing different ways of assessing the data and enabling us to draw more substantiated conclusions.

[83] Lewins, Christina, *Using software in qualitative research: A step-by-step guide*, 2007
[84] ATLAS.ti Qualitative Data Analysis, "*What is ATLAS.ti?*", 2020, https://atlasti.com/product/what-is-atlas-ti/

4 How Can We Conduct More Analytical Games Within The Constraints Of Engaging Very Senior Players?

This paper outlines a number of measures that could be employed to improve the analytical quality of national security games. However, a number of constraints potentially limit the ability of game designers to implement them in practice. Some of these constraints are somewhat universal, and extend to other forms of analysis beyond games. Other constraints are more specific to games conducted with senior participants.

4.1 General Constraints on Game Design

4.1.1 Sponsor Relationships

Good sponsor relationships are critical to the success of games. However, poor sponsors can reduce a game's quality and utility. They can do this in several ways:

- Micromanagement of game design. Sponsors often wish to take an active role in shaping the game's design. In good stakeholder relationships where final design choices are left to the game designers, this can be welcome and useful. However, in some cases, sponsors who are well-meaning but who lack detailed knowledge of gaming might insist upon design aspects which detract from the game's objectives. This might include demanding certain game mechanics, the presence of certain players, or scenario elements without understanding the implications this has on the viability of the design, or its ability to deliver the required outputs. They might also insist on a level of detail or focus in the game which is inconsistent with the study question.

- Interference with game design. More nefariously, sponsors might intervene to ensure the game advocates some pre-conceived answer[85]. Such interventions might insist on adjusting input data or adjudication models to strengthen the apparent utility of a particular option, or excluding certain inconvenient factors from the game design.

- Absentee sponsors. Senior sponsors often delegate the direction of game design to more junior subordinates until very late in the game design process. These subordinates might not fully understand the intent of the senior sponsor, leading to unclear, inconsistent and contradictory decisions and guidance. Despite not being involved in most of the design process, absentee sponsors often still reserve the right to make substantial changes to the game design at any point up to, and even during, the execution of the game, which they might feel compelled to do if their subordinates have misinterpreted the sponsor's direction[86].

4.1.2 Poorly-Defined Requirements

A lack of clear and consistent requirements is a frequent issue for game designers. Sponsors often provide requirements which are too vague and open to interpretation,

[85] Downes-Martin, "Your Boss, Players, and Sponsor: The Three Witches of War Gaming," *Naval War College Review*, 2014, Vol. 67 : No. 1 , Article 5, p.5
[86] *Ibid.*, p.3 and pp.5-6

or which are not suitable for gaming. Sponsors frequently ask for games without fully articulating what they wish to know or achieve, and why they assume – sometimes incorrectly – that a game will be the appropriate tool for this task. Stakeholders can also ask questions which are a poor fit with their wider programmes – for example asking for a System Exploration-type game at the end of a policy-making process rather than the beginning of it. At the strategic policy-making level, game designers are often confronted with constantly-changing game requirements. This can occur when game development timelines are longer than the pace at which changes occur in the policy area the game is considering. This can lead stakeholders wishing to mould the game to account for, or influence, the latest developments in the policy area in question – for example, re-shaping the game objectives to produce outputs that will influence emerging defence reviews or spending rounds. The lack of consistent objectives throughout the lifetime of a game design project makes it very hard to follow a systematic process of linking the game design to its purpose and objectives.

4.1.3 Too Much Specificity in Game Requirements

Linked to the issue of micromanagement and interference from sponsors during the game-design process is a tendency for sponsors to go beyond outlining a research question in the initial game requirement, and to instead specify details of the game design. This might include specifying a particular game method, the seniority and make-up of participants, or the setting and scenario. This can constrain the ability of game designers to subsequently make choices relating to these issues that best serve the game's analytical objectives[87]. This over-specificity relating to game design details might occur simultaneously with a lack of clarity about what the game is meant to achieve, and why.

4.1.4 Too Many, or Competing Objectives

Stakeholders frequently overload games with too many objectives. This can involve asking too many research questions, requesting that game outputs focus on too many different aspects of the problem, or requiring too many actors and relationships to be represented. This is an issue across all games, but frequently occurs where a game represents a rare opportunity to engage particular players – leading to a temptation to do everything with a single event. This is often a challenge with senior games.

Stakeholders can also specify objectives that risk competing with each other in the game design process. This is most common where stakeholders specify both 'creating knowledge' and 'conveying knowledge' objectives within a single game. Although a game designed for one of these purposes is also likely to have benefits in the other, best practice is to 'point' the game towards just one of these objectives[88]. This is because design choices taken for analytical purposes can undermine or skew the learning benefits of the game, and vice versa. For example, 'Conveying Knowledge' games might prioritise realism and immersion of players whereas 'Creating Knowledge' games might focus on less realistic, more abstract game mechanics, and be prepared to break players' immersion to collect more data from them. Despite this, sponsors frequently request that their games seek both to analyse

[87] Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, p.83
[88] DCDC, 'Wargaming Handbook', 2017, Ministry of Defence, p.9

a problem and possible responses, and to educate participants about key issues, their roles and responsibilities and so on. Whilst all of these objectives can be accommodated to a certain extent, constraints arise when sponsors demand that both analytical and learning objectives are *drivers* for the game design. For example, demanding a 'Creating Knowledge' game that is also an immersive learning experience for players.

Sponsors might request game objectives from across the typological framework outlined in section 0. Since the different typologies tend to require different game methods, and differ in the design choices that are required, objectives across typologies can require challenging trade-offs between design elements most suited to each[89]. This is most problematic when sponsors simultaneously request aspects of discovery and experimentation in the same game. This usually occurs when a sponsor wishes to build understanding of a system and rigorously test courses of action within it at the same time.

4.1.5 Compressed Development Timeframes

We are frequently asked to conduct games at relatively short notice. Where we might consider 6 months to be a reasonable timeframe in which to research, design and develop a bespoke game on a topic we had not previously considered, we are frequently given much shorter timeframes than this – in extreme circumstances ranging from a month to 6 weeks. Such timeframes significantly inhibit our ability to conduct research into the system that is the subject of the game, to work to fully understand the underlying requirements behind a sponsor's question, and to design, develop and test a game. Such time constraints can therefore lead to game design choices being driven by expediency rather than an analytical approach, and can curtail the necessary testing and refinement to ensure the game produces appropriate and valid outputs. Limited game development time also precludes testing of the game model, and evaluating the extent to which it meets the game requirement.

Often, the timeframes we are given are dictated by the need to inform a decision by a particular date. This not only imposes hard constraints on development time, but also limits the sponsor's willingness to wait for a substantial period post-game while detailed analysis is conducted. Instead, the sponsor often requires a summary report to be written immediately after the game, which is then used to inform the decision. Longer-term analysis is often either not commissioned, or not used by the sponsor. This limits the ability of the game design team to argue for thorough approaches to game design and analysis. Even where longer-term analysis is commissioned, the initial findings in the summary report often become the established narrative, even if it is subsequently refined or even contradicted by subsequent analysis. For this reason, our preference would be to decline to issue interim findings and convince sponsors to wait for the final analytical report; however we recognise that sponsors will rarely accept this.

[89] Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, p.177

4.1.6 Facilities and Technology

Many of the adjudication and data capture techniques outlined in this paper rely upon technology to enable them. Examples include computer-assisted adjudication to allow for better tracking of myriad factors and resolution with reference to underlying models, electronic capture and transmission of game board states and turn outcomes, and audio recording and computerised orders submission and surveys. However, our national security games are frequently run at classifications and in facilities that preclude the use of such technologies. It can be a major challenge to get technology accredited for use on our high classification networks. Even where such accreditation can be obtained, most senior games are run off-site, and are therefore subject to the host's security procedures. These can severely limit the use of external technology, and the addition of software to the host's networks. In the most extreme cases, the game host has even placed restrictions on the addition of PowerPoint slides on to their network. These restrictions are very likely to continue, and impose a very large constraint on the use of technology to support analysis in senior games.

4.2 Constraints Arising From Gaming with Senior Players

The majority of national security gaming conducted by the Defence Wargaming Centre (DWC) has involved senior players. Conducting games with senior players often provides invaluable insights into how senior decision-makers view and respond to national security dilemmas. Past experience has suggested that more junior players, and those with different backgrounds to the seniors, are poor at understanding the factors and perspectives that shape senior national security decision-making. Their behaviour in the game does not accurately reflect that of the seniors they are being asked to represent – as Levine states, 'game-players generally have both an imperfect knowledge of the decision-makers they are representing, and even more important, a very imperfect intuition of the way in which decision-makers feel the various pressures on them… a game tries to solve the problem of the real subtleties that govern the world (particularly, perhaps in a crisis) by using real people as play-actors – ersatz decision makers – and hoping that the real subtleties of the real people will somehow match the subtleties of the decision-makers. There is no reason to believe that they will'.[90] Where game objectives require insights into how senior decision-makers might respond to an issue, the use of senior players is therefore essential. However, conducting games with senior players imposes additional constraints on analytical game design beyond the general ones listed above.

4.2.1 Limited Game Time

The availability of senior players heavily limits the length of time available for game execution. The most we have been able to achieve in recent national security games has been two days. In some cases, we have been limited to 4 hour events. This severely restricts the ability to employ complex game mechanics, conduct thorough adjudication, properly document the game, and give participants time to reflect on the game play and offer their initial impressions. Securing senior players also requires an acceptance that higher priority tasks might cause them to only attend parts of a game – for example entire game turns have had to be scrapped in previous games because

[90] Levine, Schelling, Jones, "Crisis Games 27 Years Later: Plus C'est Déjà Vu", pp.7, 10-11

a majority of players were called to urgent real-world meetings. The combination of available time and the possible loss of players limits the ability to move a consistent cohort of senior players through an action-reaction-counteraction cycle, thus undermining the insights that such a cycle provides.

4.2.2 Gatekeepers

As mentioned above, sponsors can be absent for much of the game design process, preferring to delegate to subordinates. This is a particular issue with senior games, where the sponsors themselves are often very senior. This can create the issue of gatekeepers – people within the sponsor's command chain who limit the ability to discuss game design decisions directly with the sponsor and who make design decisions on their behalf. We have also found that subordinate stakeholders can have a tendency towards conservative design direction. They are often driven by a desire to deliver what they believe the senior sponsor expects to see, rather than enabling game designers to produce the most effective method for answering the sponsor's question. In our experience, this issue is most prominent for game mechanics. Owing to a long period in which senior decision-makers had only been exposed to structured discussions and TTXs, a conventional wisdom built up at the subordinate stakeholder level that senior players would refuse to engage with overt game mechanics, be prepared to make specific decisions, or welcome feedback on their choices. A lack of gaming awareness amongst gatekeepers can mean that they fail to see the value of these key aspects of gaming, and therefore do not wish to advocate for them with the senior sponsor.

Gatekeepers can also limit access to the senior sponsor, limiting the ability of the game design team to engage the sponsor directly on the game design choices that would best serve their needs. Written attempts to do so are often subject to staffing within the sponsor's organisation, meaning that by the time the message is received by the sponsor, it has been sanitised or misinterpreted beyond the control of the game designers. Verbal communications are often both very brief, and highly regulated by the more junior stakeholders. This is a particular issue where gatekeepers have misunderstood or misrepresented the direction they have were given by the senior sponsor. In such cases it might only become apparent at the game itself that the design and planned outputs are at variance with what the sponsor requested and expected.

4.2.3 Scepticism Regarding Game Mechanics

Recent national security games have shown that many of the concerns held by gatekeepers were not well-founded. Senior decision-makers have been prepared to make specific decisions and receive adversarial feedback on them. However, senior players have shown a degree of genuine scepticism regarding overt game mechanics. It is not possible to say with certainty why this might be, but senior players appear to be more comfortable in settings which resemble their real-world jobs and decision-making processes. They appear to become less comfortable as the issues they deal with are represented in more reductive and abstracted ways, and the roles they are asked to undertake in game differ from those that they perform in reality.

4.2.4 Limited Ability to Control the Actions of Senior Players

It can be more difficult to get senior players to complete tasks required by the game design than it is for more junior participants. Senior players might not think that the game is representative of the issues as they see them and seek to ignore or redesign game mechanics[91], they might refuse to undertake certain activities such as orders submission, player brief-back and survey population, they might become fixated on one particular issue and prefer to discuss that over continuing to play the game, or they might feel that their time is better spent on their day job and leave the game early. Whilst this issue is a possible feature of all games, it is particularly acute in senior games where there is potentially a significant gap in seniority between game facilitators and players, limiting the authority of facilitators to compel players to behave in accordance with the game design.

4.3 Options for Analytical Approaches to Games with Senior Players

Our research suggests a number of options for overcoming the constraints over analytical gaming with senior players:

4.3.1 Engage In Thorough Requirements Capture with Sponsors

It is important, very early in the game design process, to find out what the sponsor's real and most important objectives are. These are often not the same as the initially expressed game objectives: sponsors often specify a preferred solution or desired output rather than outlining a research question that gaming could contribute to the answer of. Using the guidance provided in Sections 2.5.2.2 and 2.5.2.4 of this report, sponsors can be guided towards shaping their objectives into a research question that appropriately matches their desired outputs and anticipated inputs. Following a thorough requirements capture process at the outset of the project can help clarify objectives and manage sponsor expectations about what a game can and cannot answer. Agreeing a clear and firm set of requirements early in the design process also helps reinforce in the sponsor's mind the important relationship between the requirement and the game design, and highlights the damage that later changes to the requirement or design could do to the ability of the game to answer the research question.

Stephen Downes-Martin outlines four questions that a game designer should ask at the outset of any project:

- "What do you want?" Variations of this question will often be needed to steer a sponsor away from requesting a particular type of event, and towards ascertaining the underlying issue it is hoped a game will address.

- "Why do you want it?" Initial answers to this question are likely to be very broad and cover a range of sub-topics, only some of which will be relevant to the research question. Asking supplemental "why" or "so what" questions helps the game designer focus on the particular issues and assumptions that are of importance to the sponsor and should therefore be the focus of game objectives.

[91] Downes-Martin, "Your Boss, Players, and Sponsor: The Three Witches of War Gaming," *Naval War College Review*, 2014, Vol. 67 : No. 1 , Article 5, p.4

- "Why don't you have it?" This question seeks to identify the reasons why the problem has not already been solved. This will help the game designer identify any underlying agendas behind the question, uncover analytical challenges inherent in the game objective and determine whether a game is an appropriate approach.

- "When are you rotating out of here?" Downes-Martin highlights that the interest of the sponsor's organisation in the analytical outputs of a game decline markedly after the sponsor has moved to another post. Understanding when a sponsor (and any subordinates acting as gatekeepers) are to move on helps determine the game's required timeline[92].

As part of this line of questioning, it is also important to establish what decisions or further activities the game is intended to support, and the nature of the game output that is needed to achieve this. It is also useful to clearly understand the standard of analysis a sponsor needs – will exploratory insights be sufficient or is more thorough testing of hypotheses required?[93]

The Defence Wargaming Handbook lists a further set of information requirements for analytical games, many of which should be addressed early in a project. These include the data required to support analysis, and the people whose involvement is required to ensure the game outputs are valid.[94]

4.3.2 Limit the Number of Primary Objectives and Ensure Only They Influence Game Design

During the requirements capture process, sponsors should be encouraged to limit the number and breadth of objectives they have. This will allow the game design to be focused on producing high quality information in a small number of areas, rather than spreading the game design and player's attention too thinly[95].

Where sponsors have multiple, competing objectives, especially those that might require different game designs, they should be asked to explicitly identify their primary, most important objectives. For example, where sponsors have both analytical and experiential learning objectives, they should be asked which of these is the primary purpose of the game. It should then be agreed with the sponsor that the game design will focus only on the primary objective. The achievement of secondary objectives will be pursued only to the extent that they do not undermine the primary objectives. For example, participants might be expected to derive experiential learning benefits from a game where a sponsor has highlighted analytical insights as the primary objective, but the game will not be designed to maximise experiential learning opportunities if they hinder the ability of the game to produce outputs that will answer the sponsor's key question.

Even in situations where the objectives are not ostensibly in conflict, sponsors should still be asked to prioritise if it is considered that focusing on all the sponsor's

[92] *Ibid.*, pp.7-9
[93] Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, pp.176-177
[94] DCDC, 'Wargaming Handbook', 2017, Ministry of Defence, p.53
[95] Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, p.177

objectives risk overloading the game or reducing the ability to derive useful insights. For example, sponsors might be asked to prioritise between focusing on insights relating to Blue decision-making processes or Red responses, or between a focus on strategic-level direction of a strategy and operational-level implementation of it.

4.3.3 Only Involve Seniors Where This Is Necessary For Achieving Primary Game Objectives

As already discussed, senior players can be important to answering some analytical questions. However, many of the constraints in this section can be alleviated if senior players are not involved. Where the requirements capture process has highlighted that senior players are not required to answer the sponsor's primary research question, seniors should not be invited to participate.

In many cases, experiential learning for seniors is a secondary objective of analytical games. In such cases, where seniors are not required to answer the analytical objectives, sponsors should be encouraged to consider running separate events for each objective. This could consist of initial games to analyse the sponsor's problem, and a separate game to educate senior decision-makers on the insights raised.

4.3.4 Involve Seniors as Part of a Wider Analytical Process

Where senior players are considered essential to answering primary research objectives, consideration should be given to how best to maximise the value of their inputs. In some cases, seniors might only be required to meet a subset of the research objectives. This might allow spreading the elicitation of analytical insights across a range of analytical approaches, each tailored to a particular aspect of the problem, and only some involving seniors. This would help ensure that seniors' time is focused on the aspects of the problem that they can most usefully provide input to, making best use of their limited time and increasing their buy-in to the activity. This approach helps designers avoid trying to force an entire analytical method into a very short space of time, or make major compromises in data capture and analysis to accommodate the constraints of senior engagement.

This could be achieved by treating games as individual events within a wider analytical process, rather than seeing a single game as the sole source of all the data required for analysis. This approach was applied successfully in a recent senior gaming project. Short senior Table Top Exercises were combined with working-level games and workshops to produce a complete output. An initial senior TTX focused on insights relating to UK decision-making and elicited senior-level decisions on UK strategy that were carried forward to the next stage of analysis. Working-level games, with fewer constraints on interactive and abstracted game mechanics, explored possible allied and adversary reactions to the UK's strategy. Workshops conducted in parallel focused in more detail on specific aspects of the problem. This was all combined into a set of challenges that would be presented to senior decision-makers for their response and reflection at a final senior TTX.

If a process approach is not adopted, separation of senior inputs from other analytical methods could still be achieved within a single game. For example, rather than running a two day event with senior attendance for the duration, such a game could be spread out over a longer period of time (say, a week) with seniors only attending for part of the time. One previously proposed approach suggested engaging seniors at the start of each day to seek their high-level decision-making on an issue. They

would then return to other business while this strategic direction and guidance was interpreted into an operational-level plan, which would then be subjected to an adversarial response and thorough adjudication. Conducting these steps without seniors present would allow much more time to be spent following a thorough orders submission and adjudication process, potentially allowing time to reference external models and data sources. The adjudicated result would be presented back to seniors at the start of the next day for their response and next set of moves. Such as approach would have allowed for a more systematic approach to be taken and documented, allowing for better post-game analysis, whilst still benefitting from a senior perspective. However, it would have also required the sponsor to sacrifice their secondary objectives around immersion and experiential learning for seniors, so was rejected by the sponsor.

4.3.5 Early Engagement with Senior Players

Early engagement with senior players, especially those who will be performing functions that are central to ensuring the game progresses as planned (such as Chair, Game Director and cell leads) is important to ensure that they can offer constructive criticism rather than disruption when it is too late to address their concerns[96]. Early discussions with senior players can ensure that they understand the analytical objectives of the game, how the design seeks to achieve them, and the necessary compromises that have been made to maintain the focus on the game's research goals. Engagement will help senior players understand what behaviour is expected of them, and why operating within the bounds of the game design and the roles they have been given is essential to achieving the game's objectives.

Sufficiently early engagement with seniors might uncover genuine issues that need to be addressed in the design. If a player's concerns are not well-founded, or if they appear unlikely to accept the design choices made in the game, early engagement can allow enough time to raise this issue with the sponsor, who can either intervene directly or select alternative players[97].

4.3.6 Ensure a 'Bought In' Sponsor or Senior Representative Is Present At the Game

Where the risk exists that senior players might refuse to participate in the game as intended, the presence of the sponsor, or a representative who is at least as senior as the players can be immensely useful to enforce compliance. However, this is only likely to work if the sponsor is fully supportive of the game design, and understands how the approach taken will contribute to their objectives. If not, the presence of a senior sponsor at the game might instead be counter-productive. An alternative approach could involve the use of retired senior personnel as contractors. The use of such personnel would help in two main ways:

- Use of their expertise and experience during the design phase, to help produce a game that is more likely to be accepted and engaged with by senior players;

- Use as facilitators or in other game staff roles. Retired senior personnel can help describe the benefits of engaging in the game and explain the purpose

[96] Downes-Martin, "Your Boss, Players, and Sponsor: The Three Witches of War Gaming, "*Naval War College Review*, 2014, Vol. 67 : No. 1 , Article 5, pp.4-5
[97] *Ibid.*

and importance of following certain game mechanics in language and with authority that is more likely to be accepted than if it came from more junior game personnel.

However, there are also risks to this approach. In particular, current senior players might balk at the perception that they are being directed by their predecessors. There might also be personal animosity between current and retired personnel that would need to be very carefully accounted for by the design team.

4.3.7 Prioritise Game Outputs over Perceived Realism

Analytical approaches to gaming require a rebalancing of player immersion and mechanics that generate game outputs suitable for analysis. Players' immersion and perceptions regarding the realism of the game should be considered secondary to achievement of analytical objectives. Whilst this is already standard practice at the working level, this will require a shift in the current approach to senior game design.

At present, many of the constraints outlined in this section are addressed by prioritising the Blue Cell players' perception of game realism. This helps the game make best use of senior players' expertise by placing them in a familiar environment and ensuring they feel comfortable with the nature of the decisions they are being asked to make. It also helps convince them that the game is a 'serious' endeavour and that the issues are being represented and considered in an appropriately detailed and nuanced way. However, strict adherence to representing Blue decision-making in as realistic a way as possible imposes limitations on other aspects of the game. In particular, it can substantially reduce the time and attention given to other player cells and adjudication. This risks creating an unbalanced game in which a detailed and nuanced Blue cell faces a highly abstracted adversary. It also risks curtailing adjudication to the extent that scripted or semi-scripted approaches are required, limiting the ability of opposing cells to provide truly representative feedback on player actions – see section 5. Sponsor objectives that specify that a particular decision-making body should be accurately represented also risk focusing the game at the wrong level for the achievement of the game's analytical objectives. For example, in a previous national security game, the sponsor required that MOD decision-making be accurately represented, even though the nature of the crisis being considered meant that the critical decisions would be taken at the National Security Council (NSC) level. Due to sponsor requirements, the NSC level of decision-making was highly abstracted and somewhat rushed, limiting the game's ability to produce representative outputs.

As such, in 'creating knowledge' games, sponsors and players will need to accept more deviations from reality to better serve the game's analytical objectives. Designers will have more authority to do this where it has already been agreed with the sponsor that analytical objectives are the primary objectives, and player immersion and experiential learning are secondary. It will nevertheless continue to be a challenge to convince sponsors and players that better analytical outcomes will be achieved in a game that appears less realistic. Some of the other recommendations in this paper, such as good briefing materials and senior stakeholder buy-in, will be critical to overcoming this challenge.

It should be noted that a rebalancing towards analytical objectives does not mean that senior games should abandon all attempts towards realism and immersion.

Generating useful game data requires that players behave in a manner that is representative of reality, and often requires them to replicate their real-world decision-making. Immersion can help greatly with this. As previously discussed, senior players should only be involved with analytical games because the game objectives require a senior perspective on the issues and decisions raised. If the game is so unrepresentative of reality that senior gameplay is also unrepresentative, then the value of engaging senior players at all is lost. Immersion and realism can often therefore be an important pre-requisite towards producing game outputs that are valid and suitable for analysis. However, immersion and realism should be de-prioritised where they threaten to undermine the game's analytical objectives.

4.3.8 Provision for Capturing Game Data Needs to be a Core Aspect of Game Design

In previous national security games data capture is limited, as previously described, and concerns about senior player buy-in and immersion have limited our ability to undertake more manual attempts at data capture. Since real-world decision-making processes rarely require decision-makers to systematically reflect on their thinking and the factors that are shaping it whilst they are in the midst of a crisis, this type of inquiry is also largely omitted from senior national security games. Data collection techniques like surveys and interviews have not been considered realistic or immersive and so have not been used. Since good game analysis often focuses on the discussions and thought processes that shape decisions as much as, if not more than the decisions themselves[98], failure to collect data on this has created a significant gap in the ability to analyse national security games.

As with the recommendation above regarding player immersion, greater emphasis in game design is required on how the required data will be generated and captured, even if this impinges to a certain extent on player perceptions of game realism. Technological limitations mean that data capture will continue to be predominantly a manual, and human activity. Sponsors and customers will need to accept that in complex games, capturing the data necessary to support thorough analysis will require a greater number of scribes and ethnographers than used at present, and they will need to conduct themselves in a more active, and at times intrusive way. As described above, players and sponsors will need to be persuaded that this is necessary to serve the game's analytical objectives.

4.3.9 Gaming Education for Sponsors and Stakeholders

Many of the constraints in this section arise because sponsors and stakeholders for games lack experience in commissioning, attending, and making use of the results of games. Better educated stakeholders are more likely to understand the benefits and limitations of games, the research tasks to which games can usefully be put, the role and importance of key aspects of game design, the importance of providing clear and consistent objectives in a timely manner, and the need to treat game outputs carefully. Educating sponsors and stakeholders can take many forms, and ideally should involve experience with gaming earlier in their careers and at training events.

Some activities are already ongoing to varying degrees:

[98] Longley-Brown, *Successful Professional Wargames: A Practitioner's Handbook*, pp.49-50

- More involvement of senior players in games (where appropriate). More gaming with senior players, where appropriate to the game objectives, will help socialise them to gaming, and become more accepting of game approaches. It will also help alleviate the concerns of subordinate gatekeepers regarding what seniors will accept within games.

- Better initial project briefings for sponsors. Initial meetings with sponsors usually require a brief introduction to what gaming is, how best to utilise them, and the range of methods on offer. These materials are constantly being improved and must be updated to reflect the findings in this report.

- Better general briefing materials for sponsors and stakeholders. In addition to bespoke materials used by project teams, more general publications can help sponsors understand their roles and obligations in commissioning games. The Defence Wargaming Handbook fulfils this function to a certain extent[99], but supplemental materials would help further. Annex B contains an educational note for sponsors and stakeholders which summarises some elements of this paper, providing them with an understanding of game types and guidance on how to set workable objectives.

- Reference to a greater number of examples and case studies. Being able to show prospective sponsors a greater range of analytical techniques will help them understand their options better, and understand the benefits and limitations of different approaches. Case studies of successful employment of these methods can help sponsors appreciate the benefits of following the game design team's guidance as they go through the design process. At present, strategic game designers have a comparative lack of breadth of examples to draw on. The analytical gaming research, of which this report is a part, will increase the number of case studies and proof-of-principle examples that designers can draw on in conversations with sponsors.

[99] DCDC, 'Wargaming Handbook', 2017, Ministry of Defence, p.v

5 How Can We Encourage More Representative Red Cell Responses To Blue Cell Actions?

5.1 Current Challenges in Representing Red

In order to improve the ability to generate more meaningful and valid insights, especially those pertaining to adversarial systems and the possible impact of theories of success, it is important to ensure that the adversary is adequately represented. There are, however, a number of challenges that need to be overcome to ensure that Red actors are represented in a way that is appropriate to support the generation of analytical outputs.

5.1.1 Focus On Blue Decision-Making

In national security games, sponsors generally prioritise exploring Blue decision-making and the UK and Allied processes for formulating adequate responses to crises. Because of this, and the limited game time that is often available in senior games (see section 4), the majority of available time has often been allocated to representing UK decision-making in a way that is broadly representative of reality and gives HMG players adequate time to properly consider the issues. As a result, the game mechanics and game time devoted to formulating the opposing side's response has often been curtailed in comparison. This has led to somewhat imbalanced games in which Blue moves have been the product of complex and nuanced interplay between different perspectives and preferences, while the Red Cell construct has been more simplistic. Red Cell design has tended either to assume Red is a unitary actor, or has placed heavy reliance on SMEs in the Red Cell to take account of more complex internal decision-making dynamics and constraints when formulating their moves, since these were not represented in game mechanics. This approach has been adequate where the focus has been on deriving insights relating to Blue, and can represent best practice when it is suitable to the game objectives. For example, the RAND Hedgemony game developed for the US DOD also took such an unbalanced approach to Blue and Red representation[100]. Such an approach, however, is less suitable for understanding how HMG's strategies could be perceived and responded to by a complex adversary. Often, a key objective of national security games is a deeper understanding of how the scenario and Blue strategies might influence, and be influenced by, Red internal dynamics and vulnerabilities. Simplified or 'baked in' Red responses will fail to achieve such objectives.

5.1.2 Lack of Red Cell Independence from Control

In the national security games that we have run at the senior level, non-Blue cells in the games have tended to be treated as adjuncts to the Control Cell, rather than fully independent players. The Control Cell has tended to constrain Red Cell activities to ensure that they produce actions most suited to allowing the Blue Cell to explore issues and generate insights in line with the game objectives. The extent of Control Cell oversight has ranged from suggesting or imposing certain activities or injects, to fully subsuming the Red Cell and taking over its responsibility for deciding the Red response. The inability of Red Cells to formulate independent responses to Blue actions with sole reference to their own objectives has limited their ability to fully

[100] Linick, *"Bringing Policy Games to Commercial Markets,"* Military Operations Research Society Wargame Community of Practice Brown Bag Lecture, 6 May 2020

utilise their knowledge of Red cultural perspectives and their strategic thinking and thus draw inferences from their own activities.

5.1.3 Player Objectives

Realistic representation of Red/Blue interactions can often by hindered by the provision of game starting conditions and mutually exclusive player objectives which inevitably drive players towards conflict. This can be compounded by lack of provision for players to revise their objectives as circumstances change[101]. Whilst in some real-world situations under study opposing actors will have genuinely mutually exclusive interests and objectives, in many cases the situation will often be more complex, with some of the opposing actors' interests and objectives actually being in alignment, and many of them not interacting with the other side's objectives at all. Provision of overly confrontational, and unchangeable objectives can give players little reason or opportunity to change course, and can therefore dis-incentivise efforts for negotiation, compromise or détente[102]. Escalation in games can therefore often be an inevitable consequence of the game's starting conditions, rather than free player choice. This can be problematic in games whose purpose is to study the potential escalatory or de-escalatory effects of strategies.

Overly simplistic objectives can also lead to unrepresentative adversary behaviour. 'Winning' a strategic or political game is almost always an implausible objective. Instead, positioning for the next phase of the continuous process of great-power competition is often the key to achieving a successful outcome. The core analytical focus is thus about understanding how to influence the adversary while avoiding being manipulated, while both sides avoid outcomes that lead to uncontrolled escalation. Ruinous escalation can become unavoidable if success is measured by Blue's failure alone, as Red players will inevitably be tempted to ignore the wider costs of their own actions if they are immune to consequences.

5.1.4 Lack of Inter-Cell Communication

In most of our national security games, not much provision is made for communication between Red and Blue cells. This is often because the addition of negotiations between players can add significant time to a turn structure, as allowance must be made for players to formulate an action, engage in negotiations with other players, then revise planned actions in light of the discussions they have had. Difficulties in documenting the negotiation phases of a game also create challenges for post-game analysis of why certain actions were taken, and others rejected. However, failure to provide for communications between player cells can often mean that physical, often military, actions are the only means that players have to signal other players. This can often lead to games taking on an overly militaristic, and confrontational tone for want of means to engage in dialogue and negotiation[103].

5.1.5 Red Cell Participants and Player Roles

In previous national security games, the Blue cell has usually consisted of representatives of a number of different, and potentially competing perspectives within MOD and across HMG. This allows for complex dynamics in policy formulation,

[101] DSTL/ TR116738, p.5
[102] *Ibid.*
[103] *Ibid.*

and the compromises and inconsistencies that often result, to naturally arise in gameplay. By contrast, Red and other player cells usually consist of small numbers of SMEs. Whereas in Blue Cells, where players with different perspectives represent roles and organisations with different priorities and constraints, Red cell players in national security games tend to come from within the UK intelligence community. While the views of intelligence analysts are by no means homogenous, Red cell players are rarely given specific roles or asked to represent particular perspectives within the Red decision-making apparatus. This tends to lead to non-military issues being under-represented in discussion and this can lead to an unrealistic focus on aggressive strategies. Available time, a lack of game mechanics, and the lack of player roles also tends to favour the Red Cell rapidly reaching a consensus view. In some cases this is a conscious design choice that reflects the adversary's more centralised control of government activity and simpler decision-making processes. However, the lack of representation of Red policy-making dynamics risks underrepresenting the factional interactions and constraints that affect Red decision-makers, and the constant interplay of organisational and inter-personal factors that might lead Red to making more realistic sub-optimal choices.

5.2 The Role of Red Cells in National Security Games

Improvements to the representation of Red in national security games requires consideration of what the proper role of Red should be in such games.

The 'Caffrey Triangle' reflects three different perspectives on how Red cells should play:

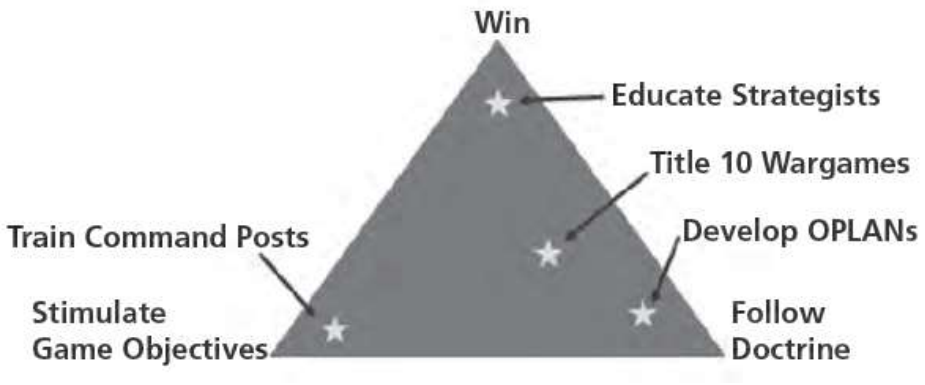


Figure 4 – The Caffrey Triangle and example applications[104]

- Follow Doctrine. Red Cells playing in this way should play as closely as possible to how Red is expected to behave in reality, strictly following Red doctrine and cultural norms.

- Win. This perspective argues that Red should play to win at all costs and provide the strongest possible challenge to Blue. They should do anything and everything a real Red actor could be capable of, without regard to factors that might constrain the Red actor in reality.

[104] Caffrey, "On Wargaming" (2019). *The Newport Papers*. 43, p.323

- Stimulate Game Objectives. In this mode, Red Cells exist solely as a tool of the Control Cell, and their purpose is to stimulate play that relates to game objectives, with little reference to what Red might do in reality[105].

Caffrey notes that proponents of each perspective often argue that only their view is correct, but in reality the way in which the Red Cell is employed should fit the game objectives. Many games will require aspects of all three approaches to varying degrees[106].

In national security games for analytical purposes, tendency towards any of the extremes of the Caffrey Triangle risks producing a Red Cell which is insufficiently nuanced and representative to produce useful outputs.

A Red Cell focused on winning in any way possible and at any cost will provide a very robust challenge to a Blue plan, which can be particularly useful in an operational game where wider strategic factors are not being considered. Traditional wargamers often suggest that Blue needs to face the most difficult challenge to be able to deal with lesser problems, but at the strategic level this understates the complexity of political discourse and often fails to address the core question the game is meant to resolve. At the national security level, 'lesser' problems are not just less serious versions of the most difficult challenge, such as a military incursion versus a full-scale invasion. Instead, 'lesser' challenges might be of a completely different nature, exploit different weaknesses and vulnerabilities and require different responses such that being able to respond to the most serious challenges is not adequate preparation to deal with the lesser ones. For example, a game that seeks to explore strategies for dealing with sub-threshold or Grey Zone threats would not be well-served by a Red Cell that immediately resorted to full-scale invasion as a worst-case response. Such an approach also risks the key players becoming disengaged. For example, an overly aggressive Red that defaults to the worst-case response regardless of what the Blue Cell does risks instilling a sense of hopelessness in the Blue players – that their efforts to engage in the game are futile since there is no prospect of them achieving an acceptable outcome. Whilst highlighting the risk of unacceptable outcomes can be a useful lesson for players, if they perceive themselves to have been unfairly punished by a Red Cell that is not subject to the constraints and pressures they would face in reality while the Blue Cell has been so constrained, they are much more likely to question the validity of the game's insights.

In national security games, freeing Red players from the need to take account of the organisational factors, constraints and costs that shape their actions risks creating misleading insights relating to Red responses to Blue strategies, especially where Blue strategies are intended to change Red perceptions of the costs and benefits of acting. This is particularly relevant to games focusing on deterrence, coercion and reassurance. For the reasons previously described, in national security games Blue cells are usually very aware of the constraints that shape their actions. If the Red Cell does not similarly take account of its constraints, the game risks becoming unbalanced, and the chances of deriving useful insights about Blue strategies, particularly those that seek to exploit or minimise the impact of Red weaknesses or fears, are diminished.

[105] *Ibid*., pp.322-323
[106] *Ibid*.

By contrast, a Red Cell that is forced to operate very strictly within the constraints of currently understood Red doctrine, culture and perspectives can be very useful for events that seek to test plans against the most likely Red response, or train players on our current understanding of Red. But a Red Cell in an analytical game that is too constrained by current assessments risks merely repeating conventional wisdom about Red, with no allowance for innovative thinking or new approaches to emerge through dynamic interaction between two cells that are both determined to gain the upper hand.

While a Red Cell that simply plays to win, or closely follows doctrine with no reference to the game objectives, might provide a robust or realistic challenge, it might also fail to produce any outputs of use to game analysts or the sponsor. The game might fail to raise the required issues, offer sufficient opportunities to explore particular policies or strategies, or stray into areas that are beyond the scope of the study. Unconstrained gameplay might also move into areas for which research has not been conducted, data has not been obtained for adjudication, or game mechanics have not been designed. However, overly constraining Red to meet game objectives risks denying them the freedom to develop what they consider to be the most appropriate responses, and the opportunity to provide useful feedback on Blue actions and strategies, thus negating the analytical benefit of gaming with interaction between live playing cells.

In our past national security games held at the senior level, Red Cells have tended towards the bottom axis of the Caffrey Triangle, balancing playing according to our current understanding of Red with providing stimulation for Blue gameplay in line with game objectives. Stimulating game objectives has often been the top priority, with playing realistically, and playing to win, being of secondary and tertiary importance respectively.

Although the precise balance between the three modes of Red play will vary depending on which game typology is being followed, we believe that Red Cells in national security games should seek a more equal balance between all three modes.

Red cell mechanics and objectives should encourage a more competitive spirit and desire to succeed, rather than merely requesting a Red ‘perspective’ based on current assessments. However, this competitive spirit should be tempered by realistic and reasonable starting conditions, objectives and constraints. This does not mean that the game design should artificially constrain Red; instead, playtests and pre-game meetings with Red should be used to explore realistic Red strategies and options that will be most effective in achieving the game’s objectives, thus allowing Red to play more freely while still meeting game objectives.

The level of Control Cell direction of the Red Cell seen in previous national security games could be reduced if some of the other game design issues were resolved. At present, very few insights can be drawn about Red’s potential responses to Blue actions because of the extent of Control Cell input to their activities. A degree of Control Cell direction to ensure that Red Cell play meets game objectives will still be necessary, however, Red Cells must be given more freedom to respond to Blue actions as they see fit. Rather than specifying particular Red actions or very narrowly constraining Red’s freedom of manoeuvre, the key dynamics the game wishes to explore should already have been identified and Control Cell inputs should therefore be more subtle ‘nudges’ designed to ensure the desired Red behaviour. Such nudges

could include the starting objectives Red is provided with, the way in which information is released to Red and the way in which interactions are adjudicated. All Control Cell interventions, no matter how subtle, should be fully documented so that they can be taken account of in subsequent analysis.

5.3 Making Red Cells More Representative

5.3.1 Red Cell Objectives

As discussed above, the Red Cell should be given objectives that encourage them to play in a manner that is situated between the corners of the Caffrey Triangle. Their objectives should include outlining their long-term strategic goals, which should always be more nuanced than simply defeating Blue. The optimum approach is to ensure that they are based upon identified Red strategic objectives and include many of the adversary's real-world perspectives, interests and constraints. This should include avoiding providing Red objectives that artificially predispose them towards confrontation with Blue. Red objectives should also encourage players to balance their responses to Blue with addressing other issues, such as managing domestic opinion or recognising the potential impact of their decisions on their relationships with other state actors. Their objectives should be grounded in as thorough an understanding of the real-world adversary as possible, but they should leave open the possibility for modification or reprioritisation to account for circumstances or opportunities arising within the game.

Such an approach might require acceptance on the part of the game designers and sponsor that the most appropriate move in a given situation might be to do nothing or de-escalate, rather than continue aggression or escalate.

5.3.2 Placement of 'Red Team' Players

It is important to note the distinction between a Red Team and a Red Cell: "the Red Cell is the enemy; the Red Team are the critical thinkers[107]".

The 'critical thinking' element of the Red Team should ideally be present in the Blue Cell advising Blue Players on potential Red policy options and likely motivations (this role often complements/supplements the Blue Cell intelligence function) while the Red Cell sits in its own dedicated room focusing on representing adversarial behaviour. Even in purely adversarial games, it is useful to have critical thinkers regularly interacting with both Red and Blue cells to ensure that the Blue players receive the information and advice they would receive in the scenario if it actually occurred.

Where political and strategic games are intended to expand the player's understanding of Red and to enhance their ability to identify both risks and opportunities, the critical thinkers can also play a vital role in providing additional background information designed to enhance or re-focus gameplay. Even in games focused on internal dynamics, the critical thinking team can be used to present an array of options to Blue decision-makers. This approach can be used to enhance or control the level of immersion created with the game environment by facilitating the presentation of complex material (including scenario updates and media commentary). The multiple perspective approach can also allow the intelligence

[107] Longley-Brown, *Successful Professional Wargames: A Practitioner's Handbook*, p.63

function to operate realistically – such as by supporting or undermining the impact of Red deception plans with appropriately contradictory reports.

5.3.3 Red and Control Cell Composition

The selection of players for the Red Team, both the 'Critical Thinkers' supporting other cells and those in the Red Cell itself, can have a profound influence on the success of a game. Critical Thinkers provide a sounding-board for Blue during game-play and are often used to assist the Control Cell. The critical thinkers don't all need to be experienced gamers but they do need to understand and prioritise the outcome that the game is intended to achieve. At least one of them needs to be able to communicate with the Red Cell. The Red Cell need to include individuals from a range of backgrounds, but there should always be a core of experienced gamers to assist control with moving the game forwards.

Red Cell players tend to come from three tribes: the first are traditional wargamers who tend to focus on 'winning the game' – a useful trait if the game is focused on exposing Blue weaknesses and highlighting the consequences of poor decisions. The second group tend to be those who prefer roleplaying games and these individuals tend to focus on character and exploring the narrative – success in the roleplayer's view is measured by the group's appreciation of the narrative and not one side's victory. In the roleplayer's view, a good narrative enhances understanding and ensures that the core objectives of the game are achieved. The third group are Subject Matter Experts, often cultural specialists or intelligence analysts. These individuals play a vital role in enhancing the plausibility of the scenario but they can sometimes lose sight of the purpose of the game and, in the most extreme cases, can assume that their limited specialist knowledge makes their opinions more valuable than others.

Red Cells (and, to a lesser extent, the critical thinkers) need to be drawn from all three of these groups, but it is important to ensure that they are in the correct roles. Roleplayers tend to make the best adjudicators as they are more likely to balance the contesting requirements created by the interaction between game objectives and the player's assumptions. They also tend to be the optimum choices for inter-factional gameplay. Wargamers make excellent Red Planners as they will eagerly explore every option to challenge Blue and exploit any weakness. SMEs need to be available to both the adjudicator and the planners but their key role is in presenting plausible options to the Blue players.

The Control Cell must also include experienced Red Team individuals. The adjudicator and umpires need to be experienced in Red Teaming and game design, and fully conversant with the objectives of the game. It is helpful if they took part in the game design process and have a thorough knowledge of the kind of scenario being played.

5.3.4 Training for Red Cell Players

The training of effective Red Teamers should balance the development of all three approaches discussed above. Red Teamers that understand the key elements of the society they are trying to represent, that recognise the importance of focusing on creating a narrative, while still maintaining the killer instinct to expose Blue assumptions and poor decision-making are priceless assets to a game project team. Appropriate training for Red Teamers should both highlight the centrality of the

game's training or analytical objectives and endeavour to blur the barriers between the three tribes. Wargamers should be encouraged to understand the context and purpose of the game, Roleplayers to use narrative processes to focus on the analytical outcome, and SMEs to use their knowledge to enhance the game.

The highly regarded US Army Red Teaming course, conducted at the University of Foreign Military and Cultural Studies at Fort Leavenworth, is costly to attend and only partially applicable for our needs (the Red Team Leader's course lasts 18 weeks). A bespoke UK focused "Red Teaming in Wargaming" course could tailor the curriculum to the types of event most often tasked to be conducted by the Defence Wargaming Centre.

Within the UK, a 'Playing Red' course was recently organised by Dstl. 'Playing Red' was a two-week instructive course intended to prepare participants to more faithfully and realistically represent a particular adversary in wargames and exercises. Several example wargames and exercises were run as part of the course.

A bespoke gaming version would shift the focus from developing an understanding of a specific adversary, with gaming as a learning tool, to further developing an understanding of the application of Red Teaming in a gaming context. Each of the gaming roles would be explored (project lead, designer, adjudicator, umpire, critical thinker, Red Cell, and ethnographer) and the development and integration of these roles enhanced by a series of brief lectures, explanatory case-studies, and training games. It is important to note that Red Teaming is never perfect – even cultural experts cannot truly represent actual Red behaviours and their decision-calculus in a context that hasn't yet arisen. What we can achieve is an improvement in the quality of Red Teaming, expand the number of trained Red Teamers, enhance the plausible representation of Red decision-making at every level, and address the issue of properly recording Red multi-faction motivations and decisions for later analysis.

5.4 Representing Red in Game Design

There are a number of ways in which game design can encourage better representation of Red. A number of suggestions are included below.

5.4.1 Communications Between Cells

Game designs should provide more opportunities for player cells to communicate with each other. There are many ways in which this could be incorporated into game designs, depending on the complexity of the structure, the need to document interactions and so on. One suggestion for allowing communications whilst keeping them manageable and recordable is to have a dedicated communications room. This can be neutral territory between the Red and Blue Cells. The points in the turn when players are allowed to enter the room, and the representatives of each cell who are allowed to engage in negotiations, might be carefully controlled. This can help mitigate the opposite issue of an overabundance of communication limiting opportunities for misunderstanding and misinterpretation, and allow players to conduct the majority of their deliberations away from the prying eyes of their opponents.

5.4.2 Provision of Information

Game designers divide games that highlight interaction into three broad types.

Open games are conducted with all players having full access to the key elements of gameplay. These are extremely useful for training games as they often involve detailed forensic discussions of the decisions and motives of both sides as the game unfolds. Openness is less useful if designers wish to explore decision-making in a more information-limited environment or wish to explore unpredictable outcomes.

Semi-closed games reveal only what could plausibly be revealed to the Blue side while including an adversarial cell which supports the game controller.

Closed games require a separate control room, often a designated communications room, and runners to update both (or more) cells on the reactions and intelligence they can perceive.

High level national security games tend to be closed or semi-closed depending on the time available, as processing decisions from both sides tends to be time-consuming. Closed games offer the opportunity to explore the decision-making of both sides based upon the limited information the scenario provides. This kind of game requires greater effort as both sides will expect the level of information they would expect in real life and this is often quite considerable. In semi-closed games, Red often has privileged access to scenario information (often assisting in its development) and this hopefully creates an opportunity for them to expand the discourse with Blue and further enhance the flow of intelligence to players. However, in the past it has often been too easy for Red to exploit this access to score points over Blue when they should really be focusing on expanding the scenario context and highlighting the risks and opportunities that Blue need to discuss. The most successful games are those that enable all of the attendees to explore the underlying dynamics of the scenario and explore the consequences of the decision-making of each of the core factions or powers. If one group dominates the game then this outcome becomes problematic. One of the key problems in all three types of game is that the players would expect more time to discuss their decisions, with the scenario often covering days, weeks, or months of time, and a combination of clarity of objectives and assertive facilitation is often required to ensure that game decisions are made and the full scenario narrative is properly explored.

The intelligence function should be there to inform players of options and not to confuse or frustrate them. Red's task is to provide the intelligence function with the ammunition to challenge Blue assumptions and inspire discourse on a range of options.

5.4.3 Observed Gameplay

A variant of semi-closed games involves providing the Red Cell with the opportunity to observe Blue's decision-making process. This approach involves the Blue Cell being observed by a dedicated Red Cell in a separate room but with a video link to the main meeting room where Blue are discussing their responses. This approach enables Red to dynamically adjust their input and maintain a focus on developing complex scenario dynamics and further enhancing the challenges that are engaging Blue. The use of specific roles for Red players and the exploitation of themed 'channels' for exchanging messages and requests for information have both been shown to enhance gameplay. The only issue with this approach is maintaining a focus on the desired outcome of the overall game as the players tend to focus on discussion rather than action.

5.4.4 Specific Roles for Red Cell Players

The Red Cell needs to reflect the key roles and dynamics that are relevant to the potential adversary. In a strategic game, it is imperative that both internal political stresses and external influences are fully represented. This can be achieved by giving each Red Cell member different objectives for the specific role they are playing. Often this will mean that defeating Blue isn't their first priority and they may even sabotage their own side to achieve their assigned objectives. This is particularly important when the Red Cell represents contending factions or groups that might resent Red success (such as criminal gangs or insurgents who do not agree with the main Red group). This hugely enhances realism but also creates fractures and tensions that Blue can attempt to exploit. Some potential adversaries have internal (sometimes chronic) tensions between services or between services and paramilitaries. A monolithic Red Cell will tend to forget these dynamics unless there is a reward for highlighting them. Red/Blue Media can also be a separate role, as they will be representing a far more complex information ecosystem than the constructed narrative favoured by the Red Cell Lead. If media were represented, the Red and Blue media representatives might both sit in the Red Cell as for many adversaries the Red Media is often broadly under Red control and Blue Media is often motivated by a desire to exploit a narrative – often one that criticises Blue.

5.5 Example Game Designs

5.5.1 Equal Treatment of Red and Blue (And Other Cells) In Game Design

One possible option to overcome previously identified issues with national security games would be to provide all cells with equal weighting from a design, game time, and resourcing standpoint, and equal freedom to make decisions as they saw fit. That is, to exactly replicate in other cells the complexity of decision-making in the Blue Cell, in terms of perspectives represented and the freedom to make independent choices. However, such an approach would require a substantial increase in game time, both to allow all the cells sufficient time to formulate actions, and also to adjudicate a complex set of activities that had not been pre-planned. Such an approach would also create challenges for participants, both in terms of identifying sufficient quantities of suitably skilled personnel to role-play in the opposing cells, and in terms of finding useful activities to occupy them in the down-time between turns while their opponents made their moves and the outcomes were adjudicated.

5.5.2 Game Focuses on Red

An alternative way of improving the representation of Red without substantially increasing the game time would be to ask the senior players to play as Red rather than Blue. Players would in effect represent their opposite numbers in the adversary's decision-making apparatus. With sufficient briefing on the objectives and perspectives of the adversary, and adequate support from adversary experts, useful insights relating to the complexities of Red decision-making, the factors governing their choices, and the constraints and pressures acting on decision-makers, could be gleaned. To simplify the game, such a Red Cell could be opposed by a substantially reduced Blue Cell, especially in situations where the HMG perspective and probable responses were already well understood. However, such an approach would only be appropriate where the game objectives were mainly focused on understanding Red

decision-making dynamics. It would also raise the issue of players representing Red being predominantly non-expert on the adversary in question.

5.5.3 Alternating Perspective

An alternative approach could be to alternate the perspective of the players by first asking them to outline a Blue option and then asking them to play the reaction of Red in the second turn. This enables the Blue players to see the situation from Red's point of view and understand how effective their initial actions might have been. This game design approach was successfully used at Standing Joint Force Headquarters (SJFHQ) in 2017 to explore options for an emerging real-world strategic challenge. During this game, it was found to be more effective to not inform Blue players (except the customer) that the game would radically change their perspective as it progressed. One of the most important aspects of this shifting-perspective approach is that it allows Blue to gain a more developed appreciation of Red concerns and vulnerabilities. The players are then returned to being the Blue cell with an enhanced understanding of the context and the Red perspective on the scenario, thus enabling the original plan to be adapted and improved.

5.5.4 Red Understanding

In the Playing Red course, the game designers used an array of game concepts to highlight the strategic and cultural insights they wanted the course attendees to understand. In the strategic/political games, the players were encouraged to develop an understanding of the motives and role of the faction they were representing and the foundations of the nation's worldview. The initial games, focused on understanding the adversary, looked at strategic prioritisation, internal political dynamics, the control of the media narrative, and the deterrence calculus. The purpose was to enhance Red gameplay and both improve the quality of in-game interactions with Blue and the plausibility of the overall scenario. Players were encouraged to consider both the strengths and vulnerabilities of the Red nation the course was exploring and to understand the key issues that might influence their future decision-making across a range of scenarios. Further games were then used to explore Red decision-making at every level from the high strategic to the tactical and in every domain. No single game could cover all of the challenges the course was intended to cover. Short and simple games designed to highlight specific issues were seen as more effective than creating an all-encompassing game.

6 Proof of Concept Escalation Dynamics Game and Concept of Analysis

This section provides an example game that has been developed as part of this research. It includes a concept of Analysis, proposes an open style of game with some codified elements as a basis for a more analytical approach for certain types of games, and puts into practice many of the lessons and recommendations we have made throughout this paper.

6.1 Deriving the Purpose of the Game

Previous work on developing analytical gaming methods for understanding deterrence suggested a number of factors that are critical to understanding deterrence but which were not well represented in Dstl's suite of national security gaming methods. One of the key areas identified was the representation of, and ability to derive meaningful insights about, escalation dynamics. The key challenges identified closely match those discussed in more detail in Sections 5 and **Error! Reference source not found.** of this paper, namely:

- Weak representation of Red, with a tendency to simplify the Red Cell in comparison to its Blue counterpart, and a Red Cell role that also included a Red team function, often at the expense of accurate representation of the adversary;
- Simplified representation of allies and third parties limiting the extent to which they can have an impact on the Blue and Red cells;
- Few means for allowing non-military Red-Blue interactions. In previous games, military actions have tended to be better represented than other levers of power.
    - In addition, the Blue and Red cells have not been permitted to engage in dialogue with each other. These have factors have tended to mean that most Red-Blue interaction has been through physical military activities and associated strategic communications, limiting players' opportunities to engage in negotiation and identify opportunities for de-escalation.

The scoping study identified three potential types of question that could be answered by a game focusing on escalation dynamics:

a. *What factors might influence the escalatory dynamics of a nuclear crisis between X and Z?*

b. *How might X deter Z and Z deter X, and why?*

c. *What opportunities could exist for de-escalation/off-ramps, and what factors would influence whether and how they were pursued?*

Of these, the first was chosen for the proof of principle game. To make it sufficiently specific to allow for an analytical design, it was re-written as follows:

"What factors might influence the escalatory dynamics of a nuclear crisis between the United States (US) and Russia?"

This is a strong System Refinement question as it focuses on more than just who the actors are and assumes some knowledge of the system – i.e. escalatory dynamics, and why this is an issue – the US and Russia are nuclear adversaries and escalation between them in a nuclear crisis is a problem worthy of examination. Breaking the question down further we can see that there are three areas of focus that we need to derive data for to answer the question:

"**What factors might influence (1)** the **escalatory dynamics (2)** of a nuclear crisis between the **US and Russia? (3)**"

6.1.1 Collection of Data: Factors

By asking "what factors might influence…" the question sets out that this is a wide based study around a subject that is not well known. When designing a game where the factors influencing a core concept are thoroughly established in the knowledge base, then the analyst should encourage the sponsor to reform the question into an Aspect or Solution type question (e.g. 'which factor (from a specific shortlist) has the most of influence on escalatory dynamics?' would be an appropriate Aspect question).

This part of the question will be the most variable, but will often be the main factor in deciding what sort of data is required for the game. For example, if the question was phrased "Which strategy would help reduce escalation between the US and Russia" then the data focus would be on the comparisons between strategies rather than on generating factors. In the example being worked through in this section, the question is looking at more qualitative and theoretical output than harder quantitative output, which can be used multiple times with different applications within the problem space process.

As a result, a number of assumptions should be included into the Concept of Analysis around this subject:

a) One play through of the game is unlikely to provide a full or even a wide range of factors. Repeat plays with the same group (for depth and also because the same players may want to try different approaches) and different groups of players (for breadth of thought) will give the best results.

b) The game will be able to provide a list of factors that *might* influence escalatory dynamics, but how much they influence the dynamics will be in rough orders of magnitude (e.g. small, medium large) and further study is likely to change these magnitudes (i.e. producing a LOW score when employing the Evidence Framework Approach). This is partially because these factors are being generated in-game, and so the adjudication of their effects will be informed by SME opinion for the most part. There is the potential that the design of the game could put a positive emphasis on some factors over others.

c) Any other evidence beyond generating potential factors to the core dynamic (escalation in this instance) will need further study to confirm their veracity.

A more free play style game would make the most sense given that the question is looking to derive factors. A codified game could be used if other aspects of the question are better known, but the question is not looking to measure or quantify anything in particular, so many of the analytical advantages given by quantified games would not necessarily be useful or analytically correct for this question. However, a hybrid system which allows players to free play decisions with some codified adjudication can be a powerful tool in these kind of games depending on the scenario and secondary questions.

6.1.2 Core Concept – Escalatory Dynamics

This question focuses on one or two actors, but there is a core concept wrapped around the actor focus. In this case it is escalatory dynamics, but it could be other concepts such as deterrence, coercion etc. Having a core concept allows designers to focus the game to include mechanics to highlight the core concept, and from an analytical perspective, designers can and should focus the data collection on the core concept. In this case the designer can now not only focus on looking for factors from a wide base, but can focus on factors specifically about escalatory dynamics, rather than just factors that affect all interactions between the US and Russia. The game may generate insights that are not about the core concept which may be valid, but it is important to recognise that the game is not focussed on these secondary considerations and so these insights may be evidentially weak.

The core concept can require a substantial amount of research time for a game as it drives the design of the underlying mechanics. A better product will be produced if the designer understands the core concept as they will have to create abstract actions and mechanics to represent it in the game. However, this is not always possible and it is often more efficient to have a study team that reports findings to the designer regarding the core concept. If this is the case, then the study team should be used for playtesting and be included in design decisions.

In this particular instance we are looking at finding factors that influence the core concept as we do not have a complete grasp on escalatory dynamics. This means that the designers would need to design mechanics that enable players to focus on the core concept and try to steer interactions to support the question. To have fully codified mechanics would therefore be inappropriate in this particular instance as fully codified games usually require a better understanding of the core concept. A more free play concept makes more analytical sense, but, some mechanics that highlight player choice consequences would be useful in the context of generating factors and when considering escalation.

For this game, this core concept means that the following assumptions and caveats must be included:

- The core concept of the game is a focus on escalatory dynamics. Any insights or factors that are identified outside of this core concept during the game may be evidentially weak and further investigation will be required before fully accepting findings on those insights or factors.

- Escalatory dynamics will be based off research and SME judgment from Dstl analysts. If future research changes our understanding of escalatory dynamics, the output of this game must be re-evaluated in case the insights

also change. (Please note, that for the purposes of this proof of principle exercise the designer's research into escalatory dynamics has been sparse; if this work was to be turned into a customer product then more research would have to be done in this area).

- The game is looking to generate factors and insights on escalatory dynamics. The findings of the game itself will improve future iterations and designs of this game. It is therefore recommended that the game is repeated with feedback from previous games being included in the designs for future iterations.

6.1.3 The Who – US and Russia

The specified focus on two actors will help generate the scenario and aid in narrowing down and simplifying the game to key actors and groups. An actor or geographical focus allows some simplifications to take place which should not affect the quality of insights. An actor focus can drive some data collection, design and analytical considerations; in this example, it is of relevance that the focus is not on the UK MOD. This means that the data that the question will drive the design to collect has a focus on escalation particularly in relation to the US – Russia dynamic. A serious consideration for Dstl when designing this game would be on how to represent the UK appropriately in this game – would the UK even be a player in this case? This would depend on the exact nature of the scenario, but over representing the UK in this instance would probably be a bad design choice.

However, if the designers were to look at this question from a UK perspective, there are probably additional insights to be learned on how the UK reacts in this situation and what opportunities and dangers there are by the UK getting involved or not. It must be recognised though that, as written, the question does not include any of these potential uses for the game. Furthermore, including them could bias the game and skew the results, particularly by giving them too much of a UK focus. The question as written is a much cleaner and analytically focussed question that would help explore escalation dynamics between two major actors. If the customers wish for a more UK focussed question, then the primary research question should be re-worded to reflect this so that when the results are published it is much clearer that this is the focus. For example, the question "How could the UK influence escalatory dynamics between the US and Russia?" is ultimately a different question to the one being examined.

For this game, the following assumptions and caveats must be included:

- The game will focus on the dynamics of escalation between the US and Russia. Other actors will be represented, but the number of actors represented will be affected by the amount of resources and time available to execute the game. This will mean that some secondary actors will gain more representation than others as a practical necessity.

As demonstrated here a strong question helps to derive the purpose of the game, some initial thoughts on data collection, and a direction in terms of general game design. If provided with a set of requirements or a set of less analytical questions, a worthwhile first step for the game designer would be to create an analytical question to answer from said requirements, which should be confirmed with the customer before proceeding.

6.2 Initial Concept of Analysis Overview: Scenario Derivation, Research and Method Proposal

We have derived the purpose of the game and identified that it requires free play mechanics to explore a breadth of potential factors and options. We have also decided that a fully codified game would be inappropriate as the purpose of the game was to increase understanding on the subject through the lens of a US-Russia crisis. Therefore, it is most likely that a Seminar or Matrix game approach is the most appropriate approach to take as a basis for the design.

However, we also know that the game will need to look at the consequences of actions, as we are looking at escalatory dynamics. This means that it would be inappropriate to use scripted narratives in the game as they reduce the primacy of the design's focus on the dynamics between different actors. Fostering a genuinely dynamic relationship between the player cells would be a key design goal in any game required to answer questions on strategic dynamics such as escalation and deterrence. These will be highly relevant when considering the game's method, but before we consider this it is important to look at scenario derivation and research.

6.2.1 Scenario Derivation

At this point in a "real" game design there would be enough information to start scenario derivation, whether this be an endorsed MOD Scenario or a bespoke scenario. For the purposes of this exercise, it is assumed that an appropriate scenario has been derived. However, the reason this step is early in the process is that it is important to identify whether an appropriate scenario to answer the question actually exists, especially if there is an imperative to use an endorsed scenario.

6.2.2 Background Research

As the purpose has been derived, it is also possible to start the research process. For this particular question the main research areas would include:

- International Relations escalation theories;
- Nuclear Deterrence theory;
- US policies and red lines;
- Russian policies and red lines;
- Other important actors.

It is important to note that background research and scenario derivation have to work hand in hand and these processes are likely to require a substantial effort which may be time consuming. Historical Analysis looking at historical examples of escalation would likely play a significant role in the background research of this game.

Since there is only one real life example of nuclear devices being used in anger – and that was to end a war, rather than start/deter one – it would be important to find analogous examples to use in the research. We must be careful not to rely too heavily on the Hiroshima/Nagasaki example as the context of nuclear use in that conflict differs significantly from the question under consideration here. We can assume that the escalatory dynamics between two nuclear armed states would vary greatly from that of a nuclear armed state and a non-nuclear state – as was the case with the Hiroshima/Nagasaki example.

For the purposes of this paper the design proposed is based on underpinning research from Lisa Carlson's *A Theory of Escalation and International Conflict*[108], where she posits the theory that escalation is based on the cost tolerances of different actors. If this game was to be used for a real question, further research would have to be done to ensure the game is representative enough of real-world dynamics.

6.2.3 Identifying Secondary Questions

The process for identifying secondary questions is an iterative one. A few secondary questions can be theoretically posited from the start as desirable outcomes of the game and as possible problem spaces the game may serendipitously answer due to its design. In this worked example, these secondary questions are:

i. What difference does knowledge of an opponent's internal pressures make to an actor's behaviour and risk of miscalculation? *This might also include consideration of the impact of imperfect knowledge or misinterpretation of knowledge.*

ii. How might pre-existing relations between the US and Russia affect escalation dynamics in the run-up to a potential conflict?

iii. What thresholds might the US and Russia see as most important in a pre-conflict scenario? Which might they be willing to compromise on?

A key consideration when looking at secondary questions is that it is reasonable to change the design of a game to help answer desired secondary questions *unless it weakens the game's ability to answer the primary question*. It is also important to recognise that the game may answer a secondary question but it may not be the best way of answering that question: a different design may be better. In this instance, these secondary questions will not affect the design of the game, but will affect data collection, as they will require more emphasis to be placed on drawing data relating to specific subject matter areas.

It is important that throughout the design process the secondary questions are revaluated when any changes to the design are made, as questions a previous design may have answered may not be answered in the final design. By extension, once the game has been play tested new secondary questions that the game may answer could be generated, and some of the desired secondary questions may need to be modified or removed. Having a list of these questions is important for future utility of the game as well, as the design could be used for other work. For more codified games it will be possible to list themes (such as Brigade Level Land Conventional Warfare, Hybrid Warfare etc.) rather than direct questions much more easily than free play games.

[108] Carlson, "A Theory of Escalation and International Conflict." *The Journal of Conflict Resolution*, vol. 39, no. 3, 1995, pp. 511–534. *JSTOR*, www.jstor.org/stable/174579

6.2.4 Proposed Method

The game method is in essence the proposed design of the game, and it should not be confused with the Analytical Method.[109] Having derived the purpose of the game, we found that the game would benefit from Free Play mechanics. However, from the core concept and the research conducted on the core concept, we can see that mechanics based on a cost calculus and consequences of actions would make the design analytically stronger. We also know from the actors identified in the question that these actions and consequences need to be made at the state level. Taking all this into account, a hybrid approach would create a strong analytical game to answer this question.

Matrix methods lend themselves to hybrid-style games more than purely discursive Seminar-style methods do as they have a more rigid structure. A matrix style game is a good approach for teasing out factors as they promote the players undertaking detailed discussions of their actions as well as rationales behind them and the reasons why they may or may not succeed.

A bespoke mechanic for the matrix game will need to be created to emphasise costs and consequences. This will focus the game around the core concept and encourage players to take actions which take these themes into account.

The concept of cost lends itself to either counters or trackers where players can spend resources. It would be easy to create a “political capital” tracker that players could spend, but according to the research this would not lend itself to the subject. According to Carlsen, escalation occurs when disparities and miscalculations in states’ cost tolerances occur. To represent this, a range of different trackers would probably need to be included that display different resources, from political to physical. Tolerances for each actor would be marked, representing the ‘red line’ limit that each resource could acceptably go down to (this information would be purposely kept secret from other actors). This mechanic would synergise well with secondary question i), as games with different starting conditions could be run – some where the tolerances are known by all the players, and some where the tolerances are hidden.

The trackers also give players a tangible resource to “attack” or “defend” with their actions, leaving the consequences of their actions less nebulous than in a pure free play game. This would mean that adjudication would need to focus on how player actions affect these trackers. Adjudication would need to take proper account of the impact of actions upon these trackers, and the consequent impact of the trackers on decision-makers, in very high stakes situations like nuclear crises. The impact of factors like Home Support and the Economy could potentially vary drastically as a crisis progresses, and the adjudication will need to be flexible enough to account for this.

Home Support	1	2	3	4	5	6	7	8	9

[109] As argued throughout this paper a game in and of itself is not analysis, a subject Peter Perla has discussed in his book *The Art of Wargaming* and at Connections UK. A game on its own is not analysis; it is the experiment designed to output data which is then analysed. Approaching game design like experimental design will therefore yield better results from games.

International Opinion	1	2	3	4	5	6	7	8	9
Economy	1	2	3	4	5	6	7	8	9
Military Power	1	2	3	4	5	6	7	8	9

Table 9 - Example of potential trackers, players would have different tolerances for these trackers

As part of this design, players would also need to give details on the ways and means of the actions they are undertaking (i.e. they should not simply state 'I want to reduce player x's home support') whilst also linking their actions to both their objectives, what they expect to achieve and what they expect the outcome to be. A risk reward matrix would help articulate how players view their actions in this respect, as Matrix Games are already good at articulating likelihood of success.

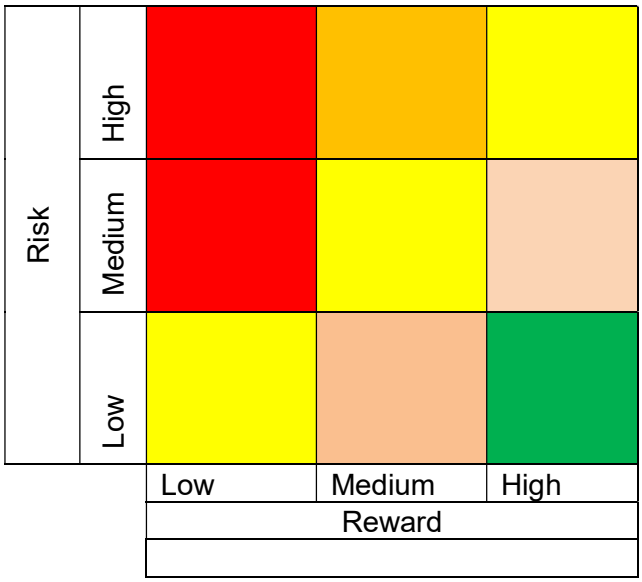


Table 10 – A Blank Risk-Reward Matrix

6.2.5 Example Risk Matrix

By getting players to articulate what they are doing with their action, what they hope to achieve and what they believe their risk and reward is from their actions, it will be easier to collect data to identify factors that may affect escalation. Crucially, other players will be able to critique their evaluation, offering their opinions on risk and reward and what the proposed actions could possibly achieve. In Matrix games a die roll often decides whether an action succeeds or fails, and if this mechanic is used then player input will help the adjudicator decide on the specific nature of the various outcomes, particularly outlier results on the dice rolls. It may be possible to directly link the Risk-Reward matrix to specific pluses and minuses to adversary or own trackers, but extensive playtesting would have to be done to achieve this.

An example Risk-Reward matrix is shown below, and we propose that it could be used in the following manner:

- The player declares their action and which tracker/trackers are being targeted;
- The opposing players would then debate which location on the risk and reward axes most appropriately represents the proposed action, thus placing the action into one of the nine boxes. The adjudicator weighs the arguments of both players before making a final determination;
- The players would then discuss what the potential outcomes associated with the four different results present within each box might look like – Exceptional Success, Success, Failure, Exceptional Failure;
- Either the adjudicator or the players would then assign probabilities to each of the four different results, and a randomisation method would then be used to decide the outcome;
- Once the outcome is decided the relevant modifier would be applied to the tracker/trackers that were selected, as shown in the table below. (For example, if a high risk and high reward action was taken which resulted in an exceptional success, then the adversary would move 4 points down on the relevant tracker).

It is worth noting that some actions – particular those that are high risk – can have negative consequences to friendly trackers even if they are successful, which represents the fact that taking high risks often has inherent negative consequences even if they pan out successfully.

Risk				
	High	Exceptional Success = Adversary -1 Success = Adversary -1, Own -1 Failure = Own -3 Exceptional Failure = Adversary +2, Own -2	Exceptional Success = Adversary -2 Success = Adversary -1 Failure = Own -2 Exceptional Failure = Adversary +2, Own -3	Exceptional Success = Adversary -4 Success = Adversary -2, Own -2 Failure = Own -3 Exceptional Failure = Adversary +2, Own -4
	Medium	Exceptional Success = Adversary -2 Success = Adversary -1 Failure = Own -1 Exceptional Failure =Own -2	Exceptional Success = Adversary -3 Success = Adversary -1, Own -1 Failure = Own -2 Exceptional Failure Adversary +2, Own -2	Exceptional Success = Adversary -4 Success = Adversary -2, Own -1 Failure = Own -3 Exceptional Failure Adversary +2, Own -2
	Low	Exceptional Success = Adversary -1 Success = Adversary -1 Failure = 0 Exceptional Failure =Own -2	Exceptional Success = Adversary -3 Success = Adversary -2, Own -1 Failure = 0 Exceptional Failure = Own -2	Exceptional Success = Adversary -4 Success = Adversary -3 Failure = 0 Exceptional Failure = Own -2
		Low	Medium	High
		Reward		

Table 11 – Example Risk Matrix with potential modifiers for the Escalation Dynamics game

To bring the focus back to escalation, there will have to be another tracker for the overall escalation during the crisis. How this tracker works would be the crux of the game, with the adjudicator and control cell deciding how player actions have affected it. This would need to involve an open discussion, allowing the players to give their feedback on how they think the tracker should move and the adjudicator explaining their final decision taking the player's thoughts into account. This will form the core part of the data collection, as scribes will need to record these conversations to identify what has caused changes in the escalation tracker. Given that escalation often arises because of differences in perception between the two sides about the situation and how it relates to each side's stakes in a conflict, it might be appropriate for each main player to have their own escalation trackers which are only visible to themselves and which they alter periodically throughout the course of the game, with the other player given fleeting opportunities to view the escalation tracker of the opponent team and potentially alter their own actions in response. This provides an opportunity for misperception and acts as a catalyst for (inadvertent) escalatory behaviour.

For this system to work the scenario would have to provide objectives to the players beyond just escalation or de-escalation. In short, there has to be reason why a player would want or even need to escalate the situation just as another player may desire to de-escalate. It is unusual in national security games to assign win conditions for players, win/loss conditions would help to drive player actions. However, these must be realistic as otherwise they may skew the output of the game considerably.

6.2.6 Data Collection

The data from this game will be qualitative in nature, focussing on finding factors that may influence escalation dynamics. This will require a scribe per team who will record the Discourse of the game whilst additional scribes record the Narrative of the Game.

The Narrative scribes will focus on recording what happens in the game, recording the events as they happened ideally with explanations as to why they happened. This will include adjudication decisions and the discussion over how the escalation tracker moves. The discourse scribes will be embedded in the teams and will be tasked with recording the reasons behind the actions players decided to take, what other options they discussed but ultimately rejected, and how they approached the problems/how the team felt about the problem. This last part is important when comparing between games; analytically it is important to know if you have a team that is enthusiastic versus a team that is devoid of ideas and is demotivated. The quality of the players is an important factor in games that can affect the overall output of the game, and only by recording this information can some comparison between play-throughs be made with different player types.

Creating a template for the scribes is important to ensure that the insights are captured in the right turn order and to prompt scribes to ask questions each turn. In some games an end of turn questionnaire may be the best way to record insights, but for much more structured problems than the one in this game design. An illustrative data capture that could be used for this game is provided in Table 12, below.

Escalation Dynamics Game	Date:	Game#
Team:	Turn:	Discourse Scribe
Actions taken this turn		
Actions Not taken and reasons		
Player insights on this turn		
Player reactions to this turn		
Player Demeanour		
Factors Identified this turn		
Primary Question Reminder:	Secondary Questions:	Questions generated during turn:

"What factors might influence the escalatory dynamics of a nuclear crisis between the US and Russia?"	A) What difference does knowledge of an opponent's internal pressures make to an actor's behaviour and risk of miscalculation? B) How might pre-existing relations between the US and Russia affect escalation dynamics in the run-up to a potential conflict? What thresholds might the US and Russia see as most important in a pre-conflict scenario? Which might they be willing to compromise on?	

Table 12 – Example scribe template for the Escalation Dynamics Game

6.3 Red Team Design and Pathways Analysis

For this game the Russian Team would be the main Red Cell, however there will be a requirement for another set of Red Teamers embedded within the adjudication cell. They would be tasked with identifying internal issues for each player nation, such as identifying that certain approaches may have detrimental economic affects to the state taking the action etc. Unusually in game design this will include a Red Teamer for the Russian Red Cell. This is not just to ensure that the traditional Red Cell are kept in check, it is also to be fair to the Russian team and provide them with the same help and guidance that the other teams get in having an SME to question their plans. Both Red Teams should be part of the Pathways Analysis and help write the Red Assumptions for the Concept of Analysis.

The justification for an expanded Red Team in the design is to help identify factors that can affect escalation. By having a Red cell focussed on adversary actions and a Red Team focussed on internal reactions and other considerations (Non-Government Organisations, non-represented countries etc.) it means there can be a greater breadth of thought and a separation of Red approaches.

This game would benefit from a Pathways Analysis to help identify Red Lines, especially from non-UK actors. This will be particularly beneficial for Red who can effectively create separate narratives using the Pathways and have a handrail to guide them based on how the narrative of the game goes.

6.4 Validation and Verification

Validation and Verification for this game will require historical scenarios to be researched and run through the game system to see if similar results can be attained. It would be impossible to get the nuclear examples to test directly, so appropriate historical case studies would be needed that can test the dynamics. Playtesting using the game scenario with external SMEs would also be required as part of verification and validation.

6.5 Example Course of Action (COA)

An example of a filled out COA is provided in Annex A.

7 Conclusions and Recommendations

This report identified a number of research questions, and below we present a number of recommendations and their supporting conclusions derived from the research we have undertaken. These have be organised according to which question they address.

7.1 How Is An Analytical Game Defined?

7.1.1 Key Definitions

Recommendation 1: The definitions and tenets outlined below should be adopted by and propagated amongst Dstl's wargaming capability.

We defined an **Analytical Game** as:

> *A game that employs analytical approaches and/or methods to generate insights as part of an analytical process.*

'Analytical approaches and methods' encompasses an analytical approach to ensuring the use of high quality inputs and appropriate mechanics during game design to engender greater confidence in insights generated, as well as the potential employment of an entire range of qualitative and quantitative methods when analysing a game to draw some form of insight.

Insights are considered to be objective conclusions drawn from post-game analysis of patterns or groups of observations made during the game's execution. We argued throughout this paper that insights from an analytical wargame should be the product of 'appropriate analytical methods' that have been applied to the data which was captured during post-game analysis, to draw conclusions that are valid and have been validated. Appropriate methods could encompass anything from within the entire range of established scientific qualitative or quantitative analytical toolsets.

We developed an analytical gaming typological framework based on work by Bartels, within which we identified a number of different types/subsections of analytical game.

In conjunction with this, section 2.3.1 distilled our definitions into a number of key tenets which underpin all analytical games:

1. **Employment of Analytical Methods to Generate Insights:** Insights from an analytical game should be to some extent the product of appropriate qualitative or quantitative analytical methods that have been applied to the data that was captured.

2. **Verification and Validation of the Game Construct:** The game's construct must be subject to a process of verification and validation to ensure it is fit for purpose, provides an accurate and appropriate representation of the real world from the perspective of its intended use, and that choices made during the design process are transparent.

3. **A Data Capture Plan:** An analytical game requires an appropriate metric collection plan which explicitly identifies what outputs from the game construct will be captured and measured, and identifies the appropriate

methods to collect them.

4. **Appropriate Data Capture:** Based on the plan appropriate data must be captured during the game's execution to provide analysts with a proper understanding of what transpired in the game.

5. **Meaningful Post-game Analysis:** Meaningful post-game analysis will be based on insights drawn from a comprehensive understanding of both 'what happened' and 'why it happened'[110] as a result of the employment of analytical methods to the data captured.

6. **Generating Novel Insights:** Post-game analysis of the game should produce insights that are not purely the product of the scenario and/or mechanisms that were an inherent part of the game's design.

7. **Generating Additional Questions:** An analytical game should also generate additional questions that will inform further research.

We would contend that all games which employ analytical methods can be defined as **Creating Knowledge** games:

> *A game that is undertaken as part of an analytical process whose purpose is gaining insights into a problem space.*

Our framework divides Creating knowledge games into two different types – **Discovery Games** and **Experimental Games**. The type of Creating Knowledge game we most frequently execute are Discovery Games, defined as:

> *A game that is undertaken as part of an analytical process whose purpose is gaining insights into an unstructured problem space.*

We consider our Discovery Games to be analytical as their inputs allow them to discover genuine high-quality insights that were previously unknown to the game design team or game sponsor, with qualitative methods of analysis being applied to the data they produce to generate said insights.

As the research around a topic matures there is the potential for us to run an Experimental Game, defined as:

> *A game that is undertaken as part of an analytical process whose purpose is gaining insights into a tightly bounded and structured problem.*

However, at present the problem spaces we operate in for national security games generally preclude us from being able to meet the criteria for a tightly bounded and

[110] Our previous experience is that the 'why it happened' element is often much more difficult to capture/understand, but is critical to undertaking credible analysis that goes beyond simplistic narrative observations of the events which took place.

structured problem from which we can produce a well-defined research question, so most of our games are likely to continue to exist in the discovery space.

7.1.2 Types of Creating Knowledge Games

Recommendation 2: From this point forward our expanded typological framework[111] should be used to characterise National Security Games, to assist the structuring of sponsor requirements and the selection of appropriate gaming and analytical methods.

Bartels proposed an archetype framework which defined a number of different types of Creating Knowledge games based on the maturity of the research and the game's purpose. Whilst we generally do not disagree with the content of Bartels's archetypes we identified an area between 'early research' and 'mature research' which we feel it did not adequately address, and that the middle ground needed to be explicitly acknowledged and described in a framework for it to have strong practical utility. This is particularly important, as we consider a significant number of our national security games and the problems we analyse to occupy this middle ground.

We would contend that at the intermediate point in the research there is a clear requirement to refine understanding/strategies that were generated during the early phases to verify and validate them and ensure that they would be amenable to more mature experimental-type approaches. Whilst this technically is touched on as part of Bartels' definitions we argue that there is a substantive difference that needs to be made explicit between games designed to promote innovative thinking and generate novel ideas in comparison to those designed to take such ideas and refine them. We therefore proposed an alternative expanded version of Bartels's archetypes:

[111] Described in detail in section 0

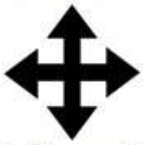
	Develop an Understanding of the System	Develop Strategies to address the problem
Early research using immature inputs which informs the research team and sponsor *(Discovery Games)*	**1) System Exploration**  Characterise and generate an initial model of the system by eliciting and synthesising current understanding	**2) Strategy Innovation**  Develop an initial 'theory of success' by generating new decision options that break from the status quo
Intermediate research using maturing inputs which evolves the understanding of the research team and sponsor *(Discovery Games)*	**3) System Refinement**  Generate a more detailed understanding of the system to refine the proposed model and increase confidence that it is a reasonable representation of reality	**4) Strategy Refinement**  Generate a detailed understanding of the strategy with a view to refining, and increasing the confidence in the efficacy of the 'theory of success'
Focused research using mature inputs which informs external stakeholder(s) in a decision process *(Experimental Games)*	**5) System Conditions**  Detect similarities and differences in decisionmaking based on different starting conditions	**6) Strategy Evaluation**  Detailed implementation of the 'theory of success' to judge the outcomes of player decisions based on a normative standard

Table 13 – The Authors' Expanded National Security Game Typological Framework

Our experiences executing previous national security games highlight that customers frequently wish to obtain outputs of evidential quality that could only come from an experimental game without understanding either, a) the sort of game that would be required to generate such outputs, or b) the quality of inputs in relation to both understanding of the problem and the theory of success that would be required to execute such a game**.**

Following this recommendation should help to address our previous experience with customers and allow us to effectively educate them as to the relationship between the purpose, inputs and outputs of games.

7.1.3 Appropriate Questions

Recommendation 3: The design team must ensure that the questions or objectives set should be appropriate to the type of game in question.

Discovery Games as a group generally require an objective or set of objectives to drive development, but it is acceptable for this not to be a specific question that

needs to be answered. Within Discovery Games, System Exploration and Strategy Innovation games occur early in the research, and so these types of games tend to have the least overall bounding in relation to their associated objectives, hypotheses or questions. System Development and Strategy Development games occur later in the research, and so their associated objectives, hypotheses or questions require somewhat more structure and focus, normally on an aspect of a wider problem.

Experimental Games involve focused research using mature inputs, and therefore require a specifically crafted question that tightly bounds the problem. Creating or identifying an appropriate question focuses the game on a specific area and allows the designer to pinpoint what data and post-game analysis will be required to answer the question, which in turn allows the designer to design the mechanics of the game to create the aforementioned data and also design an appropriate data capture plan.

There are two groups of questions experimental games should look to answer:

- Solution questions, which look at how one could achieve success and what one needs to do to achieve the 'how' (these broadly correspond to Systems Evaluation games);
- Aspect questions, which look to generate a detailed understanding of certain aspects of a problem, examining why issues occur in certain ways (these broadly correspond to Systems Conditions games).

7.2 How Can We Develop Creating Knowledge Games That Are More Analytical?

Based on our previous experience we identified a number of areas for potential improvement in relation to how we enact the gaming process detailed in sections 3.3 – 3.5.

These areas included:

- The inputs which underlie the game's model of reality;
- The development of the game's Data Collection and Management Plan (DCMP);
- Data collection;
- Post-game analysis of the data collected to generate insights.

7.2.1 A Framework for the Game Design Process

Recommendation 4: Future national security games should follow the ten step game design process outlined below.

There are a number of ways to design a game with an emphasis on analysis, the process outlined below highlights one of the best case scenarios given enough time and resources to design a game properly.

The basic process is:

I. Deriving the Purpose of the Game

The purpose of the game is the main driving force behind the game's design. The clearer the purpose of the game the more likely designers are to get a

good analytical result. To shape the game's purpose into something focused and usable to the game design team the purpose should ideally be expressed either as individual (or sets of) objectives, hypotheses, or questions that meet the overarching criteria set forward in sections 2.5.2.1 - 2.5.2.4.

II. Subject Research

This phase is about conducting research into the subject matter of the game to identify factors that influence outcomes of various actions, approaches and strategies. This step is essential to generate the baseline understanding required to abstract the real world into appropriate game mechanics.

III. Create a Concept of Analysis

After identifying an appropriate purpose, objectives, question to answer or hypothesis to test, it is important to write a Concept of Analysis. It should lay out the point of the study, the methods employed and the analysis techniques that will be used.

IV. Scenario Derivation

The game designers need to identify scenario(s) to use during gameplay. The designers do not need to create a scenario for every game and should use appropriate MOD approved scenarios if they are available.

V. Initial Game Design

This is the first attempt at designing the game and includes research into specific game designs and playtesting of different systems. Some of these processes can start before the previous phases have finished or begun, but initial design should probably not end until phases I-IV have been completed.

VI. Red Cell/Team Design

During initial game design, the integration of the Red Cell/Team into the game should be considered particularly with a focus on what role they should have during the game. Integrating some of the Red Cell/Team into the design phase is important for Red Cell/Team buy-in and to ensure that their expertise is included in the game.

VII. Pathways Analysis

This technique essentially creates a series of pathway events that could happen from a single beginning event.

This phase is non-essential as it may not apply to some game designs but can be a useful augmentation to Red Teams, and game and scenario designs.

VIII. Second Game Design period

By this period the designers should have a strong handle on the design of the game and its structure. The Red Team should be integrated and used for play testing during this time.

IX. Data Collection Techniques

This phase should be the final iterations of the data collection techniques being included in the concept of analysis and during the design phases. This phase includes production of any materials, surveys, spreadsheets etc. required for data collection and acquisition of any other recording equipment required for data collection.

X. Game Validation and Verification

The minimum requirement for validation and verification of a game is play testing and technical review. We also advocate an assessment using the EFA.

An important part of this process is returning back to the start after completing the next step to see if anything has changed and needs updating. Some of these steps can be conducted concurrently to each other, but this has to be determined on a case by case basis.

7.2.2 Codification

Recommendation 5: The rules and structure of more codified games should be treated with a higher level of scrutiny. These games should be treated in a similar manner to Dstl computer models, with log books[112] and validation/verification testing before use.

We outlined our intent to introduce analytical methods into our method for abstracting inputs into codified credible mechanics when and where possible. This would help us to create games that serve the analytical requirements put forward in the DCMP. Utilising codified mechanics can be advantageous because such games are usually more repeatable as the rules create an environment that constrains player actions to a more structured pattern. Codified games are generally easier to analyse as their outputs are easier to measure, and this can potentially allow for more quantitative analysis of the game depending on the design and whether it is appropriate to do so. We therefore believe that codification can lead to games that have more transparent and interrogable designs. However, codified games require a better understanding of the real world interactions that affect the subject matter of the game, and this may not be possible without a substantial amount of additional research.

It is also important to recognise that codifying game structures is difficult when operating at the strategic/grand strategic levels, due to the complexity and lack of structure of the problem spaces under consideration. There is an onus on the game design team to ensure that codification only takes place where appropriate, as it can introduce additional risks – for example, the employment of codified mechanics might

[112] In this context, a log book would primarily record the rationale for any codifications and log the iterations that take place as the game design progresses. As such it would not serve exactly the same purpose as a log book for a computer model.

imply we have a more definitive understanding of the problem under examination than we actually do.

7.2.3 A More Scientific Approach to Game Design and Development

7.2.3.1 Recommendation 6: Engage in thorough requirements capture with sponsors.

It is important to find out what the sponsor's real and most important objectives are, and these are often not the same as the initially expressed game objectives. Agreeing a clear and firm set of requirements early in the design process helps reinforce in the sponsor's mind the important relationship between the requirement and the game design, and highlights the damage that later changes to the requirement or design could do to the ability of the game to answer the research question. It is also important to establish what decisions or further activities the game is intended to support, and the nature of the game output that is needed to achieve this.

Recommendation 7: Limit the number of primary objectives and ensure only they influence game design.

During the requirements capture process, sponsors should be encouraged to limit the number and breadth of objectives they have. This will allow the game design to be focused on producing high quality information in a small number of areas. Where sponsors have multiple, competing objectives, especially those that might require different game designs, they should be asked to explicitly identify their primary, most important objectives, and agree that only these should be used to shape the design of the game. Secondary objectives will only be pursued to the extent that they do not undermine the achievement of the game's primary objectives.

Recommendation 8: When designing and developing games we should undertake a literature review to establish a direct link between the phenomena that will be under examination and established practical and theoretic academic understanding.

Current game design practice often involves a process of eliciting good ideas whilst using SMEs in the area of interest to advise the game team of underlying dynamics. Whilst this is a reasonable approach to take, it should be accompanied by a literature review to make explicit the links between our understanding of the problem space and the game mechanics being employed.

Recommendation 9: It is critical that all aspects of the game's design are fully documented to demonstrate how the understanding of the problem space has been applied to solve the problem at hand.

Many assumptions made by both the SMEs and the design team regarding the dynamics under study and how they are being abstracted into the game's model/mechanics are not properly documented during this process. This leads to a clearly identifiable lack of transparency in the game's design that leads to an inability to explicitly state the assumptions that were made during its creation.

Proper documentation will allow the choices made by the designers in conjunction with the SMEs to be interrogable during post-game analysis; the underlying rationales behind game mechanics will be auditable and subject to scrutiny to prove that they

were analytically sound. We can and should follow a more procedural and better documented approach to our design processes that is similar to that undertaken in computer modelling.

Recommendation 10: When we design and develop future games we need to keep in mind that analytical game design requires a conscious awareness that design choices need to be driven by objective rigour and the game's analytical requirements, not just what works for the players during gameplay.

We would also argue that our previous national security games failed to achieve a good balance between player engagement/immersion and the analytical purity of the design, often prioritising the former over the latter. Future efforts should redress this balance.

7.2.4 The Evidence Framework Approach

Recommendation 11: The EFA should be employed in line with its stated intent as a method for being more analytical with regards to our post-game analysis in national security games. At the start of the game design process the team should perform an estimate using the EFA tables on the level of evidence required from the game. As the game design process progresses, the team should schedule slots to perform evaluations based on the criteria and compare them to the initial estimates.

The EFA provides a practical way to think about evidence and improving analytical quality, and helps people become better systems thinkers by undertaking an 'analysis estimate' process. We contend that currently there are no analytical processes that we apply to analyse the game design, in contrast to the analysis applied to the outputs post-game. At present, such assessments are a product of simple expert judgement on the part of the game design team, and it could be argued that they lack structured underlying scientific procedures which would ensure objectivity.

We put forward two additional uses of the EFA which we believe will aid us in creating more analytical games, whilst allowing the EFA to act as an analytical wrapper around the entire game process from start to finish. Our contention is that with minimal reinterpretation the criteria are just as applicable to the construct and the inputs that underlie a game as they are to the outputs, and that the EFA can be used to assess the game's design as it develops. This will generate a series of EFA outputs that the designers would like the finished product to measure up to, with justifications as to why certain outputs are to be expected and should be achieved.

The results of these evaluations could either inform the design as it goes forward, and if no changes are deemed necessary they will form part of the audit trail which explains why certain decisions were made. Using the EFA in this manner will provide us with a structured framework to evaluate our game designs and help us to maintain standards of good practice.

7.2.5 Data Capture

Recommendation 12: All future national security games should include the production of a comprehensive Data Capture and Management Plan (DCMP) as part of the game design process; a proposed analytical method should be created as part of the DCMP, and this should inform the quantity and type data that needs to be captured.

One of the areas of improvement identified was the DCMP. Due to the lack of planning and clarity regarding the use of specific analytical methods the issues associated with capturing high-quality data and other constraints, we have generally not produced such a document in our previous national security games.

The DCMP is a critical document in ensuring a coherent and transparently justifiable analytical approach is maintained from the inception of the game design process to the delivery of the game's outputs to the customer.

A game needs to capture appropriate data that will be utilised as part of a considered analytical process to produce meaningful and valid insights that are relevant to the stakeholder's stated question or requirement. What data is appropriate will vary within the context of each specific game, but it will always consist of data that is of sufficient quantity and of the right type to allow the analytical methods and techniques that were pre-selected during the game's design to be properly executed.

When designing games there is a complex interrelationship between:

a. the proposed analytical method informing the quantity and type data that needs to be captured; and,
b. an assessment of the quantity and type of data the designer thinks it is realistically possible to capture informing which analytical methods will be viable.

In previous national security games we have primarily taken the latter approach, with data capture being implemented without a full consideration of analytical methods during the game's design process. This has led to a somewhat post-hoc approach to analysis based on capturing data and subsequently deciding what – if any – forms of analysis can be applied to it. In future games we need to embrace the former approach to engender a coherent analytical process throughout game design, development and execution.

Recommendation 13: Additional resources should be invested into ethnographers, to increase their numbers during execution and provide formal training.

We have already successfully employed passive methods to capture player environment data in our national security games via the use of limited numbers of ethnographers. We believe that securing additional personnel for this task alongside expanding the remit of the ethnographers would enable them to provide a substantially greater coverage of discussions during the game's execution.

Currently, it is not routine practice to train our ethnographers – we assign this task late in the process and assume that analysts already possess the relevant skills. This assumption is often faulty and leads to highly variable quality of post-game notes depending on the experience of the individual analyst in question. In conjunction with greater numbers of ethnographers, formal training would ensure that ethnographers possesses the appropriate skills to ensure high quality data collection, substantively increasing the quantity and quality of player domain data collected. This would contribute towards capturing some data which has thus far been lost. The expected value of this would be that capturing player discussions would increase our understanding of the decision making processes which led to the actions that were taken.

Recommendation 14: We should implement additional methods/mechanics in the game that could facilitate active data capture and generate different forms of data which are amenable to meaningful analysis.

Active data collection can be utilised to access types of data we have not engaged with thus far. Ordinal scales – such a likert scales, could be employed to great effect as part of an active data collection plan. In a group setting a requirement to fill out such scales could also be employed as a framing tool to facilitate discussions around topics of interest. We could also employ e-voting to ask further questions on-the-spot and address decisions/topics of interest as they take place.

Embracing active data capture methods will allow us to interview players at specific points throughout the game's execution. If ethnographers were empowered to pause the game to ask probing questions to either groups or individuals this could provide much more detailed exposure of the rationales, assumptions, and decision making processes of the players.

This would potentially allow us to access new and useful data, and enable the employment of a well-established range of social scientific methods/techniques during post-game analysis.

Recommendation 15: If a requirement is identified to use technological methods to capture data in a situation where they would bring an identified benefit, then they should not be dismissed due to practical constraints; we should attempt to engender a cultural shift in our customers so that they begin to accept such methods as routine.

A range of practical constraints have been identified have hampered data capture efforts during previous games. These include:

- time – both time required to collect data during the game and time required to subsequently analyse the data;
- reticence of senior audiences to have their conversations recorded due to strong feelings (and possible ethical concerns) regarding direct attribution of discussions;
- the amount of resource required to collect, catalogue and analyse the data in terms of manpower.

Recommendation 16: As part of the game design process a cost/benefit assessment will need to be made by the design team to recommend to the customer a reasonable amount of data capture to fulfil their requirement.

It is inevitable that the methods selected will ultimately depend on what is feasible in terms of cost. Therefore, careful consideration will need to be given to resource and the cost/benefit of different levels of data capture during the game's design so that the effect of decisions can be communicated to and understood by the customer.

7.2.6 Employment of Analytical Methods during Post-Game Analysis

Recommendation 17: In future national security games we recommend the application of analytical methods, for example content analysis, grounded theory, thematic analysis and descriptive statistics.

Our four previous senior national security games have all followed a fairly similar trajectory in terms of how the data collected was analysed to draw insights that were reported to our stakeholders. Firstly, we report the narrative that was created. Secondly, facilitated semi-structured plenary discussions during the event – generally after it is completed – are used to collect player observations into areas of relevance. Thirdly, we draw conclusions based on the narrative and player observations via the best judgement of our analysts. Finally, these conclusions are presented in a formal report and are backed up by insightful quotes from players.

Our current view is that this method allows us to produce a set of insights into the customer's problem rapidly. However, our contention that being more analytical in our approach will serve to generate more meaningful and valid insights stems at least in part from our concerns over the limitations of this process.

Throughout section 3 it was asserted that the collection of different types of data would enable post-game analytical approaches/methods that we have thus far been unable to employ. Descriptive statistics in particular could be employed to provide some degree of quantifiable assessment relating to player views on topics that are generally considered to be highly qualitative – such as escalation.

Recommendation 18: We should purchase ATLAS.ti as an enabler package to help with codifying and analysing data collected from qualitative games.

Our colleagues in the US have previously used the ATLAS.ti software package for qualitative analysis, as it can be used to help researchers uncover and systematically analyse complex phenomena hidden in unstructured data (text, multimedia, geospatial).

7.3 How Can We Conduct More Analytical National Security Games Within The Constraints Of Engaging Very Senior Players?

Recommendation 19: Only involve seniors where this is necessary for achieving primary game objectives.

Senior players can be important to answering some analytical questions. However, a significant number of game design constraints – as identified in section 4.2 – can be alleviated if senior players are not involved. Where the requirements capture process has highlighted that senior players are not required to answer the sponsor's primary research question, seniors should not be invited to participate.

7.3.1 **Recommendation 20:** Engage with senior players early in the game design process.

Early engagement with senior players is important to ensure that they can offer constructive criticism early in the game design process, rather than disruption when it is too late to address their concerns. Early discussions with senior players can ensure that they understand the analytical objectives of the game, how the design seeks to achieve them, and the necessary compromises that have been made to maintain the focus on the game's research goals. Engagement will also help senior players understand what behaviour is expected of them, and why operating within the bounds of the game design and the roles they have been given is essential to achieving the game's objectives.

7.3.2 **Recommendation 21:** Involve seniors as part of a wider analytical process.

Where senior players are considered essential to answering primary research objectives, consideration should be given to how best to maximise the value of their inputs. Games can be treated as individual events within a wider analytical process, rather than a single game being seen as the sole source of all the data required for analysis. This can allow for spreading the elicitation of analytical insights across a range of analytical approaches, each tailored to a particular aspect of the problem.

7.3.3 **Recommendation 22:** Ensure a 'bought in' sponsor or senior representative is present at the game.

Where the risk exists that senior players might refuse to participate in the game as intended, the presence of the sponsor, or a representative who is at least as senior as the players can be immensely useful to ensure compliance.

7.3.4 **Recommendation 23:** Prioritise game outputs over perceived realism.

Analytical approaches to gaming require a rebalancing of player immersion and mechanics that generate game outputs suitable for analysis. Players' immersion and perceptions regarding the realism of the game should be considered secondary to achievement of analytical objectives, although we recognise that the game should not be so unrepresentative of reality that senior gameplay is also unrepresentative.

It is likely to that will be a challenge to convince sponsors and players that better analytical outcomes will be achieved in a game that appears less realistic, but this shift in approach will be required to produce more analytical games.

7.3.5 **Recommendation 24:** Provide a degree of education on gaming for sponsors and stakeholders.

Sponsors and stakeholders for games often lack experience in commissioning, attending, and making use of the results of games. Better educated stakeholders are more likely to understand the benefits and limitations of games, the research tasks to which games can usefully be put, the role and importance of key aspects of game design, the importance of providing clear and consistent objectives in a timely manner, and the need to treat game outputs carefully.

This may involve a combination of better initial project briefings for sponsors, more involvement of senior players in games (where appropriate), better general briefing materials for sponsors and stakeholders and reference to a greater number of examples and case studies.

7.4 How Can We Encourage More Representative Red Cell Responses To Blue Cell Actions?

7.4.1 Making Red Cells More Representative

7.4.1.1 Recommendation 25: Red cell mechanics and objectives should encourage a more competitive spirit and desire to succeed, rather than merely requesting a Red 'perspective' based on current assessments. This competitive spirit should be tempered by realistic objectives and constraints; the Red Cell should be given objectives that encourage them to play in a manner that is situated between the corners of the Caffrey Triangle.

We posit that the roles of Red Cells in national security games can be illustrated via the 'Caffrey Triangle', which places Red Cells between differing modes of play – following doctrine, winning at all costs, and stimulating the game's objectives. Although the precise balance between these will vary depending on which game typology is being followed, we believe that in general Red Cells in national security games should seek a more equal balance between all three modes.

Red Cell objectives should include outlining their long-term strategic goals, which should always be more nuanced than simply defeating Blue. Their objectives should be grounded in as thorough an understanding of the real-world adversary as possible, but they should leave open the possibility for modification or reprioritisation to account for circumstances or opportunities arising within the game.

7.4.1.2 Recommendation 26: The level of Control Cell direction of the Red Cell seen in previous national security games should be reduced.

A degree of Control Cell direction to ensure that Red Cell play meets game objectives will still be necessary; however, Red Cells must be given more freedom to respond to Blue actions as they see fit. The key dynamics the game wishes to explore should already have been identified and embedded into the game's design, meaning that control Cell inputs should be more subtle 'nudges' with the intent of ensuring the desired Red behaviour. All Control Cell interventions should be fully documented so that they can be taken account of in subsequent analysis.

7.4.2 Training For Red Cell Players

Recommendation 27: Appropriate training for Red Teamers should be undertaken to highlight the centrality of the game's analytical objectives and endeavour to blur the barriers between the three 'tribes'.

Red players can be grouped into three 'tribes' – traditional wargamers, roleplayers and subject matter experts. Well trained Red Teamers – who understand their relationship to the Caffrey Triangle and can accurately represent the key elements of the society they are trying to play – recognise the importance of focusing on creating a narrative, and can still maintain the 'killer instinct' to expose Blue assumptions and poor decision-making are priceless assets to a game project team.

7.4.2.1 **Recommendation 28:** A bespoke UK focused “Red Teaming in Gaming” course should be created.

Such a course would help to further develop understanding of the application of Red Teaming in a gaming context, and provide a pool of experienced individuals to draw from when executing national security games.

8 Closing Summary

The primary purpose of this research has been to provide practical actionable recommendations relating to how we can expand our national security gaming toolset to generate more meaningful and valid insights. The research for this paper has been under way since early 2019, and its utility has already been demonstrated by its impact on the design and development of national security games we have run since its inception.

We believe that this paper should be part of an ongoing process of assessment and evaluation of the methods we use in our national security games, the purpose of which is ultimately to engender continuous improvement in the quality of the games we deliver for our customers.

We have made a significant number of recommendations covering a wide range of areas. We think that seeking to follow these recommendations will contribute to an improvement in the standard of gaming for the purpose of analysing national security problems. However, we do not intend for this document to be a definitive and final statement on the nature of national security games. Many of the recommendations in this paper have yet to be put into practice at the national security level, and we would expect that we would wish to update these recommendations and add new ones as we gather more experience and learn lessons regarding their practical utility.

We also recognise that there many gaps in this paper that would require further exploration in future work. In particular, we recognise an ongoing shortfall in tangible examples of analytical game mechanics, and entire game designs, within each of our proposed typologies. Whilst this paper seeks to address some of these gaps with practical suggestions for methods and design choices, we would hope to be able to offer more specific examples as our suite of methods expands.

We therefore recommend that this research be periodically revisited an updated to account for improvements in our own knowledge and experience, and to incorporate best practice from our colleagues in other governmental organisations in the UK and amongst our allies, industry and academia.

Bibliography

1. ATLAS.ti Qualitative Data Analysis, "*What is ATLAS.ti?*", 2020, https://atlasti.com/product/what-is-atlas-ti/
2. Bartels, "*Innovative Education: Gaming - Learning at play*", 4 Aug 2014, https://pubsonline.informs.org/do/10.1287/orms.2014.04.13/full/
3. Bartels, *Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach*, 2020, Pardee RAND Graduate School
4. Brynen, Mouat, Fisher, *Matrix Game Construction Kit: User Guide*, 2017
5. Caffrey, "On Wargaming", 2019, *The Newport Papers*. 43
6. Carlson, "A Theory of Escalation and International Conflict." *The Journal of Conflict Resolution*, vol. 39, no. 3, 1995, pp. 511–534. *JSTOR*, www.jstor.org/stable/174579
7. Collins Dictionary, 'Analytical', https://www.collinsdictionary.com/dictionary/english/analytical
8. Compton, "The Obstacles on the Road to Better Analytical Wargaming," *War on the Rocks*, 2019, https://warontherocks.com/2019/10/the-obstacles-on-the-road-to-better-analytical-wargaming/
9. Compton, Email Correspondence with Author, 14 Apr 2020
10. 'CSA's Review of Operational Analysis in Contributing to Policy, Planning and Acquisition Decision Making', July 2004.
11. DCDC, 'Wargaming Handbook', 2017, Ministry of Defence
12. "*Deductive and Inductive methods of logical reasoning*", 2012, https://bhaugolikgyan.wordpress.com/2012/12/11/deductive-and-inductive-methods-of-logical-reasoning/
13. Downes-Martin, Email Correspondence with Author, 7 Apr 2020
14. Downes-Martin, "Your Boss, Players, and Sponsor: The Three Witches of War Gaming," *Naval War College Review*, 2014, Vol. 67 : No. 1 , Article 5
15. HC264,*The Report of the Iraq Inquiry: Executive Summary*, 2016
16. HM Treasury, *The Aqua Book: Guidance On Producing Quality Analysis For Government*, March 2015, UK OFFICIAL
17. Hutton, *Review of Analysis in Defence: Interim Report*, 2015
18.

19. Levene, “Defence Reform: An independent report into the structure and management of the Ministry of Defence”, 2011

20. Levin, Fox, “*Elementary Statistics in Social Research”*, 1991, New York: HarperCollins, 6; Patton
21. Levine, Schelling, Jones, "Crisis Games 27 Years Later: Plus C'est Deja Vu," 1991, RAND Corporation, P-7719
22. Lewins, Christina, *Using software in qualitative research: A step-by-step guide*, 2007, London: Sage
23. Linick*, "Bringing Policy Games to Commercial Markets,”* Military Operations Research Society Wargame Community of Practice Brown Bag Lecture, 6 May 2020
24. Longley-Brown, *Successful Professional Wargames: A Practitioner’s Handbook,* The History of Wargaming Project, 2019
25. Lumen Learning, “*Reasoning and Inference*”, https://courses.lumenlearning.com/boundless-psychology/chapter/reasoning-and-inference/#:~:text=Abduction&text=The%20difference%20between%20abductive%20reasoning,seeks%20to%20determine%20general%20rules.
26. Macmillan Dictionary, ‘Analytical’ https://www.macmillandictionary.com/dictionary/british/analytical
27. MacGowan, “*Cataclysm: A Second World War – Rules of Play”*, 2018, https://s3-us-west-2.amazonaws.com/gmtwebsiteassets/cataclysm/CAT_Rules_FINAL_Lo-Res.pdf
28. McGrady, “Getting the Story Right about Wargaming,” *War on the Rocks,* 2019, https://warontherocks.com/2019/11/getting-the-story-right-about-wargaming/
29. McHugh, *Fundamentals of War Gaming*, US Naval War College, 3rd Edition, 1966, page 9
30. McLeod, “*Likert Scale Definition, Examples and Analysis*”, Simply Psychology, 2019 https://www.simplypsychology.org/likert-scale.html
31. Merriam-Webster Dictionary, ‘Analysis’, https://www.merriam-webster.com/dictionary/analysis
32. Myers, “Qualitative Research in Information Systems” in Patricia Y. Martin, and Barry A. Turner, “Grounded Theory and Organizational Research,” *The Journal of Applied Behavioral Science, (22:2)*, 1986, pp.141-157
33. Pearce, Email Correspondence with Author, 23 Apr 2020
34. Perla, *The Art of Wargaming,* 1990, Naval Institute Press
35. Perla, Ewell, Ma, Peachey, Sepinsky, and Tripsas, “Rolling the Iron Dice: From Analytical Wargaming to the Cycle of Research,” *War on*

the Rocks, 2019, https://warontherocks.com/2019/10/rolling-the-iron-dice-from-analytical-wargaming-to-the-cycle-of-research/
36. Potter, *An Analysis of Thinking and Research About Qualitative Methods* (Mahwah, NJ: Erlbaum, 1996)
37. Rhodes, “Friction”, 2008, http://www.strategybydesign.org/friction
38. Rittel, Webber, "*Dilemmas in a General Theory of Planning*." Policy sciences, 4(2), 1973, pp.155-169.
39. Sabin, *Simulating War: Studying Conflict Through Simulation Games*, 2014, Bloomsbury Publishing Plc
40. The University of Auckland, “Thematic Analysis: A Reflective Approach”, https://www.psych.auckland.ac.nz/en/about/thematic-analysis.html
41. Stanford Encyclopaedia of Philosophy, “*Abduction*”, 2017, https://plato.stanford.edu/entries/abduction/
42. Terdoslavich, Muldoon, “*Cataclysm: A Second World War – Playbook”*, 2018, https://s3-us-west-2.amazonaws.com/gmtwebsiteassets/cataclysm/CAT_Playbook_FINAL-Lo-Res.pdf
43. Wasser, Oberholtzer, Pettyjohn, Mackenzie, “*Gaming Gray Zone Tactics: Design Considerations for a Structured Strategic Game*”, RAND Corporation, 2019, https://www.rand.org/pubs/research_reports/RR2915.html

List of abbreviations

DCMP	Data Collection and Management Plan
DBOWS	Divisional/Brigade Overlay Wargame System
CAT	Confidence Assessment Table
COA	Course of Action
DIME	Diplomatic, Informational, Military and Economic
DoD	Department of Defense
Dstl	Defence Science and Technology Laboratory
DWC	Defence Wargaming Centre
EFA	Evidence Framework Approach
EPT	Evidence Profile Table
HMG	Her Majesty's Government
HSA	Hostile State Activity
LI	Lessons Identified
MOD	Ministry of Defence
NATO	North Atlantic Treaty Organisation
NSA	National Security Advisor
NSC	National Security Council
OILs	Observations, Insights and Lessons
RAF	Royal Air Force
SME	Subject Matter Expert
SJFHQ	Standing Joint Force Headquarters
TTPs	Tactics, Techniques and Procedures
TTX	Table Top Exercise
UK	United Kingdom
US	United States
VPT	Validation Profile Table
VCDS	Vice-Chief of the Defence Staff

Annex A: Analytical Game Activity COA

Game: Escalation Dynamics Game

COA Author: Stephen Ho

Activity Lead: Stephen Ho

Date:07 April 2020

Section (a)	Sub-Section (b)	Details (c)
Introduction	Primary and Secondary Questions	The primary question for this study is: “What factors might influence the escalatory dynamics of a nuclear crisis between the US and Russia?” The purpose of this study is therefore to investigate potential factors that may influence escalatory dynamics through the lens of US-Russian relations during a nuclear crisis. Secondary questions that may be answered by this game include: A) What difference does knowledge of an opponent’s internal pressures make to an actor’s behaviour and risk of miscalculation? B) How might pre-existing relations between the US and Russia affect escalation dynamics in the run-up to a potential conflict? C) What thresholds might the US and Russia see as most important in a pre-conflict scenario? Which might they be willing to compromise on? A manual game approach has been chosen due to the complex nature of the subject matter and the flexibility by manual games.
	Exploitation (Explains how the output/evidence will be used (assisting the revision of doctrine, capability planning, supporting IG/MG business cases).	The output will be used to generate insights and recommendations about escalation dynamics for UK MoD and wider gov, identifying factors that will be of use in similar crises.

	Level of Evidence Required The levels of evidence required will depend on what the evidence is being used for, insights from a single activity may be good enough in some cases (e.g. writing concepts and doctrine), however supporting equipment business cases that require Scrutiny approval will usually require statistically robust evidence from multiple activities (See Annex A FDA&E Handbook on Levels of Evidence)	As the game is an exploratory game, levels of evidence required are relatively low as breadth is more important than breadth. Multiple play-throughs should also be conducted to increase this breadth. **Evidence Profile estimates** **Comprehensiveness: 3 Due to the complex problem space of International Relations and Escalation Dynamics that not all key aspects and associated uncertainties will be fully explored**. This game should be considered the first step for understanding Escalation Dynamics through games and is looking to find factors but not quantify them. Further investigation of these factors will be required afterwards. Comprehensiveness could increase if the game is repeated enough times. **Relevance: 2** The sources of data for Escalation Dynamics theory is diverse and offers many different perspectives. However, many disagree with each other and none of the sources are directly relevant to the question at hand. Historical, real life examples are relatively rare. **Objectivity: 2** Evidence used to base the dynamics of the game on will come from peer reviewed, respected journals and publications. However, the way it has been implemented in the game will be from Dstl interpretation as such a method has not been conducted before, bringing the score down to 2. **Quantity: 3** The game will form the sole source of data for which the conclusions (insights) will be drawn, but the game itself will draw on a wide range of SMEs as well using qualitative methods of analysis. **Consistency: 4:** Cause and effect are not well understood, in this field and there are multiple perspectives on the problem.

	Governance (Lists the key customers/stakeholders and how the study will be governed. A wiring diagram is often used.)	Customer = Dstl Study Lead = James Bennet Manual Wargaming Lead = Stevie Ho Stakeholders for further exploitation: Dstl = Mike Bagwell
	Timescales & Deliverables (Provides the critical dates when output is required and the form in which it should be provided (report/presentation). - Key decision/activity (e.g. WFE periods) points - Deliverables (a table with firm dates)	TBD
	Scope - Focus – details the key entities/systems being assessed (e.g. BG, LE TacCiS) - Timeframe(s) (e.g. 2025, 2030)	Timeframe = 2020 to 2025. The focus of the game will be on determining factors for Escalation Dynamics in a US-Russia Crisis. The scale will therefore be Grand Strategic with players representing nation states. All levers of government should be represented in the game, no special focus should be given to the military.
	Scenarios Describes why certain scenarios have been chosen. Is a scenario characterisation required? Consider type (intervention, stabilisation), environment, terrain (wooded, mountainous and urban), distances to travel to theatre/objective, threat level, coalition contribution.	Bespoke
	Blue and Red Orbats/ Capabilities	N/A

	Assumptions Lists the key assumptions (e.g. Air/AH might need to be excluded for experimental purposes so that ground elements can be appropriately explored).	d) One play through of the game will potentially not provide a full or even a wide range of factors. Multiple play throughs with the same group (for depth and also because the same players may want to try different approaches) and different groups of players (for breadth of thought) will give the best results. e) The game will be able to provide you with a list of factors that MIGHT influence escalatory dynamics, but how much they influence the dynamics will be in rough orders of magnitude (e.g. small, medium large) and further study is likely to change these magnitudes (ie a LOW score on Paul Pearce's Evidence Framework). This is partially because these factors are being generated in game and so the adjudication of their effects will be informed by SME opinion for the most part and there is the potential that the design of the game could put a positive emphasis on some factors over others. f) Any other evidence beyond generating potential factors to the core dynamic (escalation in this instance) will need further study to confirm their veracity. g) The core concept of the game is a focus on Escalatory Dynamics. Any insights or factors that are identified outside of this core concept during the game may be evidentially weak and further investigation will be required before fully accepting findings on those insights or factors. h) Escalatory Dynamics will be based off research and SME judgment from Dstl analysts. If future research changes our understanding of Escalatory Dynamics, then the output of this game must be re-evaluated in case the insights also change. (Please note, that for the purposes of this exercise the research into Escalatory Dynamics has been sparse, if this work was to be turned into a customer product then more research would have to be done in this area). i) The game is looking to generate factors and insights on Escalatory Dynamics. The findings of the game itself will improve future iterations and designs of this game. It is therefore recommended that the game is repeated with feedback from previous games being included in the designs for future iterations.

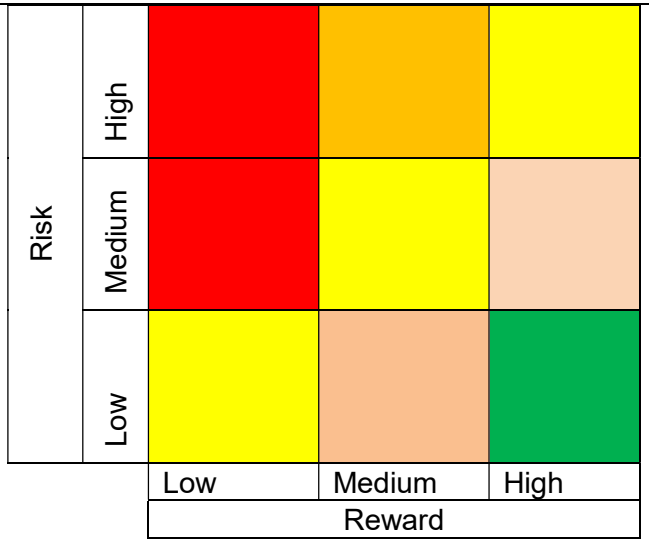
		j) The game will focus on the dynamics of escalation between the US and Russia. Other actors will be represented, but the number of actors represented will be affected by the amount of resources and time available to execute the game. This will mean that some secondary actors will gain more representation than others as a practical necessity.
Analysis Approach	Intent	Analysis from the game will be mostly qualitative. They will draw on insights from the outcomes of the games and the consequences of player decisions.

	Method Describes the hypotheses/options/cases and/or variations being assessed. Details the tools that will be used to conduct the analysis, and how they will be used together to assess the question/options/cases under examination	A matrix game with codified rules modifications. A specific mechanic for the matrix game will need to be created to emphasise costs and consequences, focussing the game around the core concept and encouraging players to take actions that take these themes into account. To represent this, trackers will be used to represent different resources (political or physical) with nation tolerances to the levels each resource could go down to being hidden from other players. The trackers also give players a tangible resource to “attack” or “defend” with their actions, leaving the consequences of their actions less nebulous than in a pure free play game. This would mean that adjudication would need to focus on how player actions affect these trackers.

Home Support	1	2	3	4	5	6	7	8	9
International Opinion	1	2	3	4	5	6	7	8	9
Economy	1	2	3	4	5	6	7	8	9
Military Power	1	2	3	4	5	6	7	8	9

Example of potential trackers, players would have different tolerances for these trackers

As part of this design, players would also need to give details on what actions they are doing (ie they should not be able to say, I want to reduce Player x’s home support without saying what they are doing to try and reduce that support) whilst also linking what actions they do take to what they actually want to achieve and expect to happen. A risk reward matrix would help articulate how players view their actions in this respect, as Matrix Games are already good at articulating likelihood of success.



Example Risk Matrix

By getting players to articulate what they are doing with their action, what they hope to achieve and what they believe their risk and reward is from their actions, it will be easier to collect data to identify factors that may affect escalation. Crucially, other players will be able to critique their evaluation, offering their opinions on risk and reward and what their actions could possibly achieve. In these games a die roll often decides whether an action succeeds or fails. If this mechanic is used then the risk reward mechanic will also help the adjudicator decide on what happens on fails and successes, particularly outlier results on the dice rolls. It may be possible to directly link the Risk Reward matrix to specific pluses and minuses to adversary or own trackers, but extensive playtesting would have to be done to achieve this.

An escalation tracker would be used to track overall escalation during the crisis and Matrix arguments around this tracker will be an important data collect for the game.

	Run Plan and Sensitivity Analysis Lists the cases (baseline and treatments/variations) that will be examined.	Several Baseline games would have to be conducted to get the most breadth. A second series of games could include allowing other players to know the tracker boundaries of each other nation's trackers to see if knowledge of each other's internal tolerances changes decision making (one of the secondary questions). This might include devising mechanisms for varying the degree to which players have imperfect knowledge of each other's internal pressures, or allowing for a degree of misinterpretation of the information they have to hand.
	Measures of Merit (MoM) Describes the MoMs e.g. Measures of Effectiveness/Performance (MoEs/MoPs) that are to be used.	Nations will be given objectives to help drive the game, but it may also help measure degrees of success from the game. Narrative and Discourse assessments of the game will form the qualitative insights from the base game and the effects of the variations.
	Costs Is the analysis to include costs e.g. cost the options under examination so they their cost-effectiveness can be compared.	N/A
	Output It is often useful to provide an example of the desired output format (tables, graphs, etc.) so the customer knows what to expect.	Report

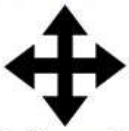
	Validation of Tools and Data If the required tools and data are not ready to use what development is required? It is also important to know the validation status of the tools and data that will be used.	The main validation system for the rules will be playtesting them. i.e. running through and seeing if they work and give a reasonable result in the time allocated). Playtesting the system through historical examples will also be important. Log book to be developed prior to the event and updated following the event.
	Links to other Studies/ Experiments Describes the interaction that will take place with other relevant studies (e.g. the provision of data)	
Analysis Plan	Tasks & Timescales Lists the key activities and when they are.	
	Resources States who is going to undertake the key activities?	
	Output/Deliverables Provides a comprehensive description of the outputs of the study, as the introduction will only list the primary ones.	Deliverables = Insights and factors Game Rules
	Risks Describes the major risks and how are they being mitigated.	Subject too complex to game effectively

Annex B: Sponsor Handout – National Security Game Archetypes and Objectives for Sponsor Awareness

This note outlines the range of game types that can be conducted to consider national security issues. The purpose of this note is to help game sponsors refine their objectives and expectations to ensure that they match the available game inputs and required game outputs. It will help sponsors work with game designers to identify the most appropriate game archetype and establish an achievable set of primary objectives. The archetypes and definitions have been adapted from one proposed by a RAND analyst, with additions and alterations to match Dstl experience[113]. A full explanation of the rationale for the archetypes in this note can be found in DSTL/TR123926.

National Security Game Archetypes

Strategic national security games fall into six broad types, as illustrated below:

	Develop an Understanding of the System	Develop Strategies to address the problem
Early research using immature inputs which informs the research team and sponsor *(Discovery Games)*	**1) System Exploration**  Characterise and generate an initial model of the system by eliciting and synthesising current understanding	**2) Strategy Innovation**  Develop an initial 'theory of success' by generating new decision options that break from the status quo
Intermediate research using maturing inputs which evolves the understanding of the research team and sponsor *(Discovery Games)*	**3) System Refinement**  Generate a more detailed understanding of the system to refine the proposed model and increase confidence that it is a reasonable representation of reality	**4) Strategy Refinement**  Generate a detailed understanding of the strategy with a view to refining, and increasing the confidence in the efficacy of the 'theory of success'
Focused research using mature inputs which informs external stakeholder(s) in a decision process *(Experimental Games)*	**5) System Conditions**  Detect similarities and differences in decisionmaking based on different starting conditions	**6) Strategy Evaluation**  Detailed implementation of the 'theory of success' to judge the outcomes of player decisions based on a normative standard

[113] Bartels, Elizabeth M., Building Better Games for National Security Policy Analysis: Towards a Social Scientific Approach. Author, 2020. https://www.rand.org/pubs/rgs_dissertations/RGSD437.html

The games in the centre column are designed to build an understanding of a problem-space, or system, while those on the right are designed to develop and test strategies, or theories of success for responding to that system. Some degree of understanding of the system is necessary before strategies for operating within it can be explored. The rows in the table above represent increasing levels of maturity of inputs to the game, in terms of the level of understanding of the system, or level of detail in the strategies to be evaluated.

Each of the six game types outlined above require different methodologies, use different inputs, and generate different types of output.

Games in the first row tend to correspond to the early stages of research. They are usually initial attempts to characterise a system or come up with ideas for effective strategies within it. The outputs of these games are often broad insights and characterisations of problems, and initial ideas regarding solutions that would require later refinement.

- **System Exploration** games are designed to elicit participants' perspectives to establish a foundational characterisation of a problem. The output of such games will be an initial model of a system and broad insights relating to its salience, potential boundaries and important aspects.

- **Strategy Innovation** games are designed to encourage innovative thinking about potential responses to a problem. They are intended to produce candidate solutions, or 'theories of success' that would be refined and tested further as understanding of the problem, and the proposed solutions, matures. A defining feature of Strategy Innovation games is that players are tasked with coming up with a strategy in the game itself; however, games in this archetype can range from those which output very initial 'blue sky' thinking about broad approaches through to which develop a detailed and coherent 'theory of success'. Such games will almost always require a broad characterisation of the system as an input. Games which are intended to help produce detailed strategies as outputs might also require broad strategies or guiding strategic principles[114] as inputs.

Games in the second row represent the intermediate step between initial ideas generation and very mature understanding and strategies. Games in this row are used to improve models of the problem and refine strategies to the point where they could be subjected to more detailed experimentation and evaluation. Such games require a working model of the problem or a candidate strategy, or strategies, as inputs to the game. These could be derived from games in the first row or other sources. The intent of these games is to refine these inputs by subjecting them to challenge, with the understanding that if problems or weaknesses are revealed, the model or strategy is still subject to change and iteration. These games – often in conjunction with other research – challenge the model or 'theory of success' that was proposed in the early stages of research to ensure that the key underlying assumptions hold together. Game players are provided with a model or strategy of

[114] These might include, amongst other things, descriptions of factors deemed to be particularly relevant to a strategy, key capabilities a strategy is expected to exploit, or key perceptions, vulnerabilities or behaviours a strategy should seek to target.

sufficient maturity that they are able to focus more fully on operating within a system or implementing a strategy, rather than initially characterising a system or inventing strategies.

- **System Refinement** games build on the outputs of early research with the intention of generating a comprehensive model of the problem system. Post-game analysis of game outputs will seek to build towards a model that the designers consider to be a reasonable representation of reality. The model generated as a result of the game, and post-game analysis, will be suitable for potential employment in other areas of research such as forming a baseline for System Conditions games or analysis and as a context for detailed strategy development. To enable these outputs, System Refinement games will require a pre-existing characterisation of the system as an input..

- **Strategy Refinement** games build on the outputs of early research with the intention of building towards a comprehensive and fully refined 'theory of success' that has been subject to challenge, thereby providing reasonable levels of confidence in its efficacy. The focus of these types of games is on iterative improvements to a strategy through adversarial challenge. Players will be tasked with implementing a well-defined strategy rather than developing one. The outputs of such games will range from identification of areas where strategy improvements are required, through to a strategy or 'theory of success' which is sufficiently mature and of a level of detail that it is suitable for exacting testing and evaluation.

Games in the third row seek to produce more robust insights relating to mature systems models and fully formed and refined strategies. Games in the third row are more experimental in nature. To produce rigorous outputs, they tend to focus on testing specific hypotheses or evaluating bounded aspects of a strategy. These games are not designed to make on-the-fly improvements to systems models or refinements to strategies. Instead, they take more detailed models and strategies as inputs and seek to produce outputs relating to the impact of particular variables on a model, or evaluate the effectiveness of particular aspects of a strategy.

.

- **System Conditions** games seek to understand how a key factor shapes decision-making processes and choices. They explore the impact of different starting conditions on the system and the choices faced by decision-makers within it. These games are often heavily bounded[115] to isolate the impact of a particular variable.

- **Strategy Evaluation** games seek to evaluate policies and strategies. They provide information about the outcomes of a proposed solution with sufficient fidelity to allow the utility of a plan to be judged.

[115] Given the complexity of the systems in question in some cases these boundaries will be artificially imposed by the game designers for the purposes of creating a workable game framework. Ideally these boundaries would also be informed by the outputs of a previous system refinement game or other analysis.

Setting appropriate primary objectives

Because of the different inputs, outputs and methodologies employed within each of the six game types outlined above, in general it is highly recommended to ensure all game objectives occupy just one box within the typology. This is because objectives that occupy multiple parts of the typology will tend to create tensions between game design choices that are ideally suited to each. At the very least, this risks creating a very complex game. At the worst, there is a danger that the use of game mechanics which are suited to one set of objectives contradicts or undermines another set of objectives. For example, it is very unlikely that a game designed to generate strategies will also be able to test them. This is because innovation games need to encourage creativity from the players and be open to a vast range of possible player actions. By contrast, evaluation games are built with a particular strategy in mind, and only those factors and actions which are directly relevant to its implementation will be included in the design, to ensure external variables do not unduly complicate or skew the analysis.

Even within single boxes of the typology, care must be taken to ensure that objectives are not too wide-ranging and unfocused. For example, objectives which seek to develop strategic and operational aspects of a plan in the same game risk over-complicating it, even though they both occupy the 'Strategy Refinement' typology. This is because the players, adjudication expertise and data inputs that are required for each level are different, and the interdependence of strategic and operational factors risks creating a complex feedback loop between players at each level. Ideally, one of these levels of planning should become a fixed assumption or an independent variable which is provided to players as an input. If both levels of planning are to be the subject of player activity, then other objectives should be de-prioritised to ensure the game design is focused on producing high quality outputs in a limited number of areas, rather than low quality outputs in a larger number of areas. In this example, that might require de-emphasising Red and third party responses to the Blue strategy.

Where such narrowing-down of objectives is not possible or desirable, particular objectives should be nominated as primary objectives, whilst others should be considered secondary. Only primary objectives will drive the game design, while achievement of secondary objectives will be sought only to the extent that doing so does not undermine achievement of the primary objectives.

All games will require objectives of sufficient clarity and specificity to allow the most appropriate methodology to be selected to match the game's intended purpose. It is always insufficient to ask for a game without being able to articulate an intended purpose. Objectives should always be set with an awareness of the probable maturity of game inputs, a good understanding of the quality of game outputs required, and a plan for how the expected outputs will be used. However, the precise nature of the required objectives varies significantly with game type.

Examples of questions for each game type

System Exploration and Strategy Innovation games are expected to occur early in the research process, and are intended to be exploratory. It is therefore expected that these types of games will tend to have the least overall bounding in relation to their

associated objectives, hypotheses or questions. Example questions and sub-questions could include:

- What is the nature of the problem?
    - What vulnerabilities do we have in domain [X]?
    - What threat might [adversary X] pose in context [Y]?
    - What might the implications be of problem [X] for us?
- What should we do about problem [X]?
    - Does this problem require us to respond?
    - What should our objectives be in deciding how to respond?
    - What should our strategy/policy/plan be to address this problem?

Since System Refinement and Strategy Refinement games are expected to produce more precise outputs than Exploration and Innovation games, their associated objectives should also be more structured and focused. As they occur later in a research process, it is expected that existing system knowledge or nascent strategies will support the development of more bounded objectives. These might seek to focus on exploring or identifying particular aspects of a wider problem, identifying implications in a particular area of an issue or strategy, or subjecting specific, defined strategies to scrutiny and challenge. Example questions could include:

- How might pre-existing relations between actor [X] and actor [Y] affect escalation dynamics in the run-up to a potential crisis?
- What factors might lead to conflict between [X] and [Y] in situation [Z]?
- What challenges might we encounter as we go about implementing this policy?
- We capabilities do we need to deliver this strategy?

System Conditions and Strategy Evaluation games are intended to generate high quality evidence about the operation of specific factors within a system or the likely impact of particular aspects of a strategy. They are run along similar lines to scientific experiments, isolating specific variables for careful study. Because they are designed to produce very focused outputs, they require specifically formulated questions which tightly bound the problem. Since System Conditions and Strategy Evaluation games occur at the mature stages of research and strategy formulation, it is expected that sufficient knowledge of the system, or detail in a ‘theory of success’ will exist to allow the development of very focused objectives. System Conditions and Strategy Evaluation questions should address either a specific variable or specific aspect of the system, and should be written in such a way that they could theoretically be given a binary yes/no answer (even though this is extremely unlikely to be the case in practice). If the question is broader than this then it would probably not be considered to constitute an experiment.

Systems Conditions games tend to focus on *Aspect Questions*, which seek to generate a detailed understanding of certain aspects of a problem, examining why issues occur in certain ways. The focus for Aspect Questions is on testing and enhancing one’s understanding of how particular factors and issues operate within a well-defined problem space. Examples of appropriate Aspect Questions include:

- Do clear red lines help with deterrence strategies?

- Is escalation necessary for successful coercion?
- Does knowledge of an opponent's internal pressures change an actor's behaviour and risk of miscalculation?

Strategy Evaluation games tend to focus on *Solution Questions*, which examine what needs to be done for a strategy or approach to achieve success. The priority for Solution Questions is in testing options available to the players and exploring a specific aspect of the problem space through a lens of pre-identified potential solutions to problems. Examples of appropriate Solution Questions include:

- Does [theory of success X] allow the UK to achieve conventional deterrence against [peer adversary Y] in [scenario Z]?
- Would [strategy X] allow the UK to meet all of its HADR commitments if the rate of natural disasters increased?

Initial distribution

1.	Catherine Nichols	DST	electronic
2.	Lucy Evans	DST	electronic
3.	Christine Wilson	DST	electronic
4.	Keith Goodman	DST	electronic
5.	Laura Dod	DST	electronic
6.	Alan Lawrence	Dstl	electronic
7.	Mahesh Vaisnav	Dstl	electronic
8.	Becci Baser	Dstl	electronic
9.	Capt Matt Harvey	Dstl	electronic
10.	Col David Batty	Dstl	electronic
11.	Gp Capt Chris Platt	Dstl	electronic
12.	KIS	Dstl	electronic

Report documentation page v5.0

* Denotes a mandatory field

1a.	**Report number: *** DSTL/PUB131779		1b.	**Version number:** 1.4
2	**Date of publication: *** 28/05/2021		3.	**Number of pages:** ii + 144
4a.	**Report UK protective marking: ***	UK OFFICIAL		
4b.	**Report national caveats: ***	NONE		
4c.	**Report descriptor: ***	NONE		
5a.	**Title: *** How Can Dstl Expand Our National Security Gaming Toolset To Generate More Meaningful And Reliable Insights?			
5b.	**Title UK protective marking: ***	UK OFFICIAL		
5c.	**Title national caveats: ***	NONE		
5d.	**Title descriptor: ***	NONE		
6.	**Authors: *** Bennett, J.; Bagwell, M.; Ho, S.; Strong, P.			
7a.	**Abstract: ***			
7b.	**Abstract UK protective marking: ***	UK OFFICIAL		
7c.	**Abstract national caveats: ***	NONE		
7d.	**Abstract descriptor: ***	NONE		

While Dstl has a lengthy track-record of delivering senior Table Top Exercises (TTXs), recent national security games were the first time in many years that dynamic, interactive games on national security issues had been conducted in the United Kingdom (UK) with serving senior personnel. The games we have conducted so far have added considerable value in identifying the challenges the UK faces and driving action to address them. Whilst we believe that the games we have conducted so far have added value in understanding blue processes, tools and capabilities, they have fallen short in areas such as engendering truly strategic dynamics, understanding possible Red responses to Blue actions, and how responses to crises might play out.

There is a growing interest in the Ministry of Defence (MOD) and across Government in the use of gaming to explore strategic issues in complex wicked/unstructured problem spaces. There is a recognition amongst sponsors and stakeholders that gaming could be used to provide insights into strategic questions where existing empirical and deductive methods have been found wanting. To fully address the questions that are being asked of us by senior Her Majesty's Government (HMG) stakeholders we have identified a requirement to generate more meaningful and reliable insights from our national security games.

Given the above the authors assert that increasing the level of challenge and analysis in Dstl's national security games will serve to generate more meaningful and reliable insights, and the intent of this report will therefore be to provide some potential solutions as to how this can be accomplished within the constraints under which our national security games are required to operate.

8.	**Keywords:**

Please note: Unclassified, Restricted and Confidential markings can only be used where the report is prepared on behalf of an international defence organisation and the appropriate prefix (e.g. NATO) included in the marking.

* Denotes a mandatory field

9.	**Name and address of publisher: ***	**10.**	**Name and address of funding source:**
	Dstl Exploration Division Portsdown West Portsdown Hill Road Fareham PO17 6AD		MOD Chief Scientific Advisor MOD Main Building, Whitehall, London, SW1A 2HB
11.	**Funding source contract:**		
12.	**Dstl project number:**	709513	
13.	**Programme:**	Policy and Capability Enterprise Support	
14.	**Other report numbers:**	DSTL/TR124295 (Summary Report)	
15a.	**Contract start date:**	**15b.**	**Contract end date:**
16.	**IP conditions for report: ***	MOD Funded Dstl authored reportMOD Funded Dstl authored	
17a.	**Patents:**	NO	
17b.	**Application number:**		
18.	**Release authority role:**		

Guidance on completing the report documentation page can be found on the Gov.UK website.

THIS PAGE INTENTIONALLY LEFT BLANK