Defence Academy Yearbook
2014

A Vibrant, World-Leading Academy
The Defence Academy Yearbook

Academic year 2013/14

A selection of commended essays from courses delivered by the Defence Academy

With a foreword by Vice Admiral D L Potts CB

Director General Joint Force Development and

Defence Academy

© Crown Copyright and individual authors

Editor: Jane Volpi
Communications Co-ordinator, Defence Academy
PUBLISHER’S NOTE:

The views expressed in this Yearbook are entirely and solely those of the individual authors and do not necessarily reflect official thinking and policy of either Her Majesty’s Government or of the Ministry of Defence.

EDITOR’S NOTE:

There have been minor alterations to the texts of these papers to conform with the constraints of this volume.

© All Serving Officers’ work is published under Crown Copyright.
© Crown Copyright, individual authors and Cranfield University.
Contents

Foreword by Vice Admiral D L Potts CB Director General Joint Force Development and Defence Academy................................................................. 1

Historical Evaluations

Between hopes and disappointments: Evaluating the Church during the Great War
Hanneke Takken .................................................................................................................................................................................. 2

What was the reality for the average Frenchman under occupation and how did the Germans exert control of the population.
Gp Capt D Toriati ................................................................................................................................................................................. 10

What factors explain the Italian participation in multinational military operations over the last decade?
Maj D Paduano .................................................................................................................................................................................. 18

The German forcing of the Meuse was the ‘inevitable outcome of the advantage the attacker has in choosing his point of attack’
Col J S A Carr-Smith ......................................................................................................................................................................... 40

As the experience of the French and German forces in 1940 demonstrates, doctrine, not equipment, is the key to success in modern warfare.
Col S Scheibl ................................................................................................................................................................................... 49

The Blame Game: An analysis of the outbreak of war in 1914
Lt Col Maynard ............................................................................................................................................................................ 57

Assess the impact on the Allied war effort of Germany’s V-weapons offensive.
Gp Capt M Johnson ............................................................................................................................................................................ 78

Addressing Today’s Defence and Security Challenges

A critical examination of the role of leadership in the handling of National Security Crises
Sqn Ldr S Gray ...................................................................................................................................................................................... 86

The strategic dynamics of international collaboration for defence procurement and support
Gp Capt S Ellard .................................................................................................................................................................................. 91

Deterring non-state actors: A dialogue of perspectives
Lt Col M Ackrill ............................................................................................................................................................................... 105

Complexity, war, and the complexity of war
Lt Col D Bickers .............................................................................................................................................................................. 126
The International Dimension

What are the impediments to the creation of an independent Kurdish state?
Lt Col J Ainley .................................................................................................................. 150

China’s peaceful rise or otherwise: Implications for security and prosperity in the Asia-Pacific
Col C Websdane .............................................................................................................. 173

Shale: A fracking geopolitical game changer
Capt I Bisson .................................................................................................................. 192

Is the Black Sea an area of future conflict or cooperation and what are the implications for regional maritime security?
Lt Cdr T Davey .............................................................................................................. 207

How do we end piracy on the high seas?
Sqn Ldr M Lock .............................................................................................................. 228

The Moral Component

Courage: Cardinal virtue or core value?
Rev Alan Steele .............................................................................................................. 249

Religion in the public arena: A strategic challenge facing the Church of England
Rev Jonathan Chaffey ................................................................................................... 256
Foreword

Vice Admiral Duncan L Potts CB
Director General Joint Force Development and Defence Academy

I am pleased to introduce this selection of papers produced by students studying at the component colleges of the Defence Academy over the 2013/14 academic year (September 2013 – July 2014).

The Defence Academy of the United Kingdom delivers postgraduate education and the majority of command, staff, leadership, defence management, acquisition and technology training for members of the UK Armed Forces and MOD Civil Servants. The broad portfolio of Defence Academy activities is intended to enhance the understanding, skills and competences of our Service and civilian personnel so that they are able to respond swiftly and imaginatively to the challenges of an increasingly uncertain world.

In the academic year 2013/14, the Defence Academy delivered over 384,000 days of face-to-face training and education to more than 22,000 students. This total included around 1,300 international students* from some 80 countries, who attended courses delivered both in the UK and overseas. During the course of their studies, Defence Academy students also undertook various forms of blended learning that amounted to over 32,000 equivalent training days. Wider education activities include a lunchtime Seminar Programme run at MOD Main Building, DE&S Abbey Wood, HQ Army and HQ Navy Command.

The Defence Academy’s teaching is delivered through colleges and centres across the UK. The Academy works in conjunction with its academic providers, Cranfield University and King’s College London.

The papers have been brigaded into 4 broad categories: ‘Historical Evaluations’, ‘Addressing Today’s Defence and Security Challenges’, ‘The International Dimension’ and ‘The Moral Component’ and represent only a sample of the range of subjects addressed by students over the academic year. Restricted space means that we can only offer a small sample selection of papers in the Year Book; however a much wider selection is available for reference on our website - www.defenceacademy.mod.uk.

January 2015

*statistical data for student numbers is attendances not individuals
Between hopes and disappointments: Evaluating the Church during the Great War

Hanneke Takken

Introduction

The First World War was not only a military conflict. It pulled whole societies in; men, women, children, workers, politicians, artists, intellectuals. In their own way, the churches became very much involved as well. Many leading churchmen chose to support their nation’s war effort. Churches in participating nations did their bit by opening their buildings at all hours to give people the opportunity to pray for their loved ones at the front, by collecting money for war funds, calling their flock to economise in the use of fuel, food and clothes when these became scarce. But they did not only give material support. As the traditional spiritual guide of the nation, many churches strongly felt the responsibility of giving meaning to war. Many a churchman spoke or wrote of war as a punishment, a purification or as a mission from God. Many defended their nation’s actions, justified their war as a defensive one and emphasised their people’s responsibility to have faith and to persevere. The German churches have often been accused for their supportive role during the war. Last July, both the German Catholic Church and the German Protestant churches officially declared themselves partly responsibility for the war enthusiasm of 1914. The German Protestant Church state that their theology had failed to meet up to their responsibility of striving for peace and reconciliation. The churches’ belief in a loving and reconciling God, their bounds with other churches through Christ’s body and the universality of their faith ‘did not save them in 1914 from war enthusiasm and war propaganda, nor for the justification of national war goals up till the very end. (...) Today, we are overwhelmed with deep shame for this failure and this guilt. We need and we want to learn from this. How can we prevent such errors in the future?’

Hopes and disappointments

The questions the German Protestant Church asks today seem very similar to the questions that the European churches asked themselves during and directly after the war. There had been an opportunity for the churches in 1914 to act as moral guidance in a world breaking down in war. Initially, many churchmen believed they were doing just that. However, as war progressed, an increasing sense of failure and a searching for meaning is clearly present, especially in German and British sources. An example of this is the conversation between the Archbishop of Canterbury Randall Davidson and Herbert Henry Asquith – Prime Minister until 1916 and leader of the liberal party – in early February 1918. During the war, the two men had regular meetings, discussing war and politics. Davidson relates this particular conversation in his diary:

I asked him [Asquith], did he think that the Church, using that word in its widest sense, was fairly chargeable with having failed to find and use its opportunities during the war. He replied that on the whole he did think so, though it was difficult to speak positively, and he admitted that everyone is apt to criticise everyone else at such a time. I asked him to put his finger, if he could, on any special lack of duty on our part, or to name any time during the war when it had been specific and definite. He could not do so, and I pressed him as to whether the dissatisfaction in the matter, which undoubtedly exists, is not the rather fretful criticism of other people which is the outcome of general dissatisfaction with the condition of things. It would be different if he could name anything specific, but he certainly could not.

When war broke out, it had been met with certain expectations. Ideas of the church’s task in society and in the field, guiding the nation as a whole and the soldiers in the field. After a while, it
became clear that many of these expectations could not be met. War increasingly seemed to raise questions more than it provided answers; on the position of the churches in their societies before the outbreak of the conflict, on the part the church and its clergy was to play during the war, on the future role and responsibilities the church was to take in the post-war world. The conversation between Davidson and Asquith shows how difficult it was to formulate answers on these questions. But it also shows a clear awareness that the church had failed to raise up to the opportunity war had seemed to offer and the need to rethink the church’s role in society.

This article will treat the hopes and disappointments that can be found in three important publications reviewing the church’s war effort: the two-volume study of the German Protestant Church’s war effort by Martin Schian (1921-1923), the British report on religion in the army by David Cairns (1919) and the evaluation of the work of the French Catholic military chaplains by Geoffroy de Grandmaison and François Veuillot (1923). These sources are different in many respects: their authors, their motives for writing, their central question and scope. But they all share the urge to extract practical lessons from the experience of war and they all regard the war as the agent of change – for better or worse.

In the summer of 1916, the German Protestant Church Commission (the confederation of German Protestant state churches) asked theologian Martin Schian to write a thorough report on the church’s war effort. This resulted in two volumes, one on the church’s war effort at home (published in 1923) and one concentrating on the front (published in 1921). The report is based on official documents of the German chaplaincies and collected personal accounts from chaplains. When Schian received his commission in 1916, his mission was to counter the critique the church increasingly received during the war and to justify the good work chaplains were doing at the front. However, its publication was overhauled by the German war loss. The book received a new agenda; to establish ‘the causes for the gradual moral exhaustion of the army during the war and to justify the good work chaplains were doing at the front. However, its publication was overhauled by the German war loss. The book received a new agenda; to establish ‘the causes for the gradual moral exhaustion of the army during the war leading to the complete internal and external collapse’.4

D.S. Cairns’ report The army and religion. An enquiry and its bearing upon the religious life of the nation was first published in 1919. This project focussed on the development of religiosity at the front. Accordingly, it used evidence from many groups that had worked with soldiers: chaplains, but also medical personnel, privates, hut leaders and workers. The initiators – a committee of theology professors, ministers and bishops from the Church of England, the Church of Scotland and one reverend from the Central Baptist Church of Bloomsbury – aimed to ascertain ‘what was being revealed under War conditions as to the religious life of the nation, and bring the result before the Churches.’5

The two French Catholic laymen Geoffroy de Grandmaison and François Veuillot had coordinated the distribution of voluntary military chaplains during the war. In 1923, they published a book reviewing the work of the chaplains and soldier-priests in the army and the role of their chaplains bureau in supporting them. It is mainly based on the correspondence the bureau had had with chaplains during the war and contains many witness accounts by military chaplains, discussing their work with the soldiers and the problems and opportunities they encountered during their work.

These sources seek to draw lessons from the war; lessons regarding the organisation of the chaplaincy, the relation between church and state and the church’s role in society and the heritage of the shortcomings of the church before the war. All three publications, though in different degrees, show an interesting movement from viewing war as a perfect opportunity for the church to realise a national religious revival toward recognising war as a learning school for the church and clergy instead.
Initially, many churches had hoped war would bring religious revival, both at home and at the front. A well known and well researched perception that can be recognised in France, Britain and Germany. All the signs of a religious revival seemed to be there. People found their way to church again. Churches were packed during those first weeks of war. Schian writes that ‘Never has the church been less criticised as during that time, never is so thankfully taken what she had to give.’\(^6\) Especially the military chaplains were thought to be essential to help this hoped-for religious revival on its way. The front seemed a new missionary field, where thousands of men would seek comfort and support in the chaplains’ words – men who had been increasingly hard to reach at peace time. However, as war progressed, the religious fire that had initially blazed so radiantly seemed slowly to die down. As Grandmaison and Veuillot cite one chaplain saying: ‘The results are ok for those troops who are under fire, mediocre or nil for those who are out of the reach of the shells.’\(^7\) An observation which is shared by Cairns and Schian. Cairns concluded that ‘If we are basing our hopes for religion in the future on the idea that the men are going to set the Churches on fire with a new zeal for religion, we shall assuredly be disappointed’.\(^8\)

Promising as the times of war had seemed, war had not brought the national religious revival the churches had hoped for – not at home, and not at the front. What had gone wrong? The publications discuss a number of factors influencing the church’s war effort among the soldiers. Two related issues stand out: firstly, the problems regarding the organisation of the chaplaincies and the position of the chaplains within the army and secondly, the susceptibility of the potential army flock.

**A troubled organisation**

All three chaplaincies – the British, the French and the German – strongly developed during the war. However, they all suffered from problems regarding an insufficient number of chaplains in the field and confusion about their exact duties, responsibilities, working area and distribution. The organisation of British and German chaplains was closely aligned to the military hierarchical structure. There were higher and lower ranking chaplains. The Army’s Chaplains Department was responsible for the British chaplains. At the eve of the Great War, one chaplain was attached to each brigade, or one to every three or four regiments.\(^9\) Garrisons and barracks at home were to be served by local clergy. There were four classes of commissioned military chaplains ranking Captain (4th Class), Major (3rd Class), Lieutenant-Colonel (3rd Class) and Colonel (4th Class), their rank depending on years of service. In charge of the regular chaplains was the Chaplain-General, who was himself submitted to the Secretary of War.

The German chaplaincy knew two organisational models: the South-German model and the Prussian model, which was dominant. The South-German model gave civilian churches more authority in military church affairs compared to the Prussian model, which was more strictly military and hierarchical. The chaplains were full members of the military and, as such, submitted to and part of military hierarchy. Military commanders were by law obliged to help the chaplain to execute his tasks, by for instance accommodating time and space for religious services and making sure that soldiers had the opportunity to attend.\(^10\) One Protestant and one Catholic Feldpropst were responsible for the chaplains of their own creed. They were appointed by the monarch and submitted to the ministries of War and of Religion. As such, the Feldpropst was strongly aligned to the military and state leaders. The Protestant Feldpropst did officially have a seat in the Protestant Church Board, but the German civilian church had hardly any influence on the Feldpropst. The Feldpropst was assisted by the *Militäroberpfarrer*. One was attached to every army corps and stationed at the High Command. Every army division was served by two divisional chaplains. These were appointed by the Feldpropst.
While the British and German chaplaincies were well integrated in the army’s organisational structure, the French chaplaincy was not. As a matter of fact, there were military chaplains, but no proper organisation. In the Gallican tradition, French royalty and bishops have always refused to accept an hierarchical organisation of their military chaplains under the auspices of the Pope. Individual bishops did not want to sacrifice their own authority over their clergy for a central organisation headed by a episcopus castrensis – possibly under the direct responsibility of the Parisian archbishop Léon Adolphe Amette. Additionally, the anticlerical government and the consequential separation of church and state in 1905, led to a very difficult relation between church and state which had direct consequences for the position and work of the military chaplains in the French army. The church was not welcome to mingle in the army of the secular republic, which made it difficult to initiate an official chaplaincy for the French troops.

Notwithstanding anticlerical measures, there were official Catholic chaplains attached to the French army – if only a small minority of the mobilized French clergy in the army. The French Catholic clergy can be divided into three separate groups: those ordained between 1889 and 1905 served as stretcher-bearers and at medical units in case of war. Those ordained after 1905 were submitted to conscription and served accordingly as regular soldiers in the troops. And finally, there was the group of official military chaplains. Since 1913, four chaplains were attached to each army corps: each group brancardiers de corps d’armée (BGC) was issued two Catholic, one Protestant and one Rabbi chaplain. Additionally, each group brancardiers de division (GBD) received one Catholic priest (ministre catholique). These official chaplains, aumôniers titulaires, received the military title and pay of captain and were responsible for religious services in the field – if necessary assisted by local priests. They were soon assisted by 250 extra voluntary chaplains (aumôniers volontaires), who were mobilised on the initiative of the Catholic layman Albert de Mun with the consent of prime minister René Viviani. These voluntary chaplains were coordinated by Geoffroy de Grandmaison and François Veuillot through the Bureau des aumôniers in Paris after De Mun’s death on 6 October 1914.

Additionally, a third group of unofficial chaplains joined the ranks, the aumôniers bénévoles. These were priests serving as ordinary soldiers but who performed religious duties within their division, often at the explicit request of their military superiors. Some became fulltime unofficial military chaplains who received financial support from their home dioceses or the Chaplains’ Bureau to buy the things they needed to execute their new tasks, such as portable chapels, prayer books and chalices. For others, being a fighting soldier remained their core business, organising religious services and individual support when time and opportunity presented themselves.

Schian and Grandmaison and Veuillot point at organisational problems regarding the church in the field. The French organisation of the chaplains suffered from three main problems: an insufficient number of chaplains, a bad distribution of those chaplains available and a bad administration which made it difficult to register the positions of the chaplains within the army. The majority of official chaplains were stationed at the ambulances. This not only resulted in having problems reaching the soldiers in the distant firing line, but it also contributed to the chaplains’ reputation of being cowards, avoiding the danger the men were exposed to. The policy to keep chaplains at the ambulances was based on the conviction that chaplains were only there to tend to the mortally wounded, not to serve the healthy soldiers. The authors point at the separation between state and church being the core of the problem. This led to the secularisation of state, army and society and had resulted in the reduction of the chaplain to someone only responsible for the sick, wounded and dying. Grandmaison and Veuillot write that here was a ‘profound ignorance of the spiritual needs of the men, of the real role of military chaplains during campaigns and of what the needs of a religious organisation are.’
On top of that, there was no good-functioning organisation to assist these French chaplains and defend their interests. Grandmaison and Veuillot emphasise that this lack of ecclesiastical organisation of the chaplaincy placed the chaplains at the mercy of doctors and military commanders. Regulations were often unclear or even contradictory. According to the authors, military commanders made good use of the lack of central organisation of the chaplains for their own benefit. For instance, they appointed soldier-priests as chaplains rather than use the official chaplain, as they had direct military authority over their own soldier-priests.13

Like Grandmaison and Veuillot, Martin Schian sees the relation between church and state as an important explanation for problems the chaplains encountered during their work. While the French chaplains were too much on their own, their German colleagues were too dependent on the Obrichkeit, the ruling elite. Schian is very explicit about this. He recognizes that it would have been impossible to realise a fully independent chaplaincy within the army, but he strongly emphasised that the highest commanders, and especially the emperor, could and should have given these chaplains more liberty and independence. He states that ‘in retrospect, they [the chaplains] were in their religious service really much restrained’. 14 This was not only the result of the tradition to make everything and everyone subordinate to the military, states Schian, it was also grown from the willingness of the church bodies to adapt to the vision and opinion of the state, which had developed in the relation between church and state through centenaries. However, Schian does not fail to add the rather bold assertion that a complete respect for religious and moral points of view would and should have led to painful conflicts.15

A troubled flock

The second and related issue concerns problems regarding the susceptibility of the chaplains’ potential flock. The authors, especially Cairns, point at problems reaching the men at the front. In order to get the chaplains' message through, relations needed to be good between the chaplain and his flock, the soldiers needed to be open for his message and the message itself had to be clear and straightforward, without the hint of a hidden agenda. That this was often not the case is abundantly described in these three sources.

Although Grandmaison and Veuillot repeatedly emphasise the good relations between chaplains and soldiers and officers, they also report clear manifestations of anticlericalism in the army. They cite chaplains’ accounts on all kinds of harassment, such as the prohibition of wearing the soutane, the removal of the chaplains’ horses, the refusal to deploy military musicians for the religious service and the ban on wearing religious symbols on military uniforms and French flags. Several chaplains testify that they are confronted with the so called rumeur infâme. These encompass wide-spread rumours that the French Catholic Church had wanted war as revenge for the separation between state and church and hoped for a German victory. One chaplain even claims to have been threatened to be shot on the accusation of being a German spy.16

From these three publications, Cairns’ report describes the distance that existed between the church and the men at the front most extensively. As Cairns argues, the church had lost its grip on the men. Several factors had caused this. The church had failed to speak a language the ordinary man, the soldier, could understand. Moreover, it had failed to convince these men of the importance and relevance of the church’s message. Because of poor religious education, the soldiers had difficulty understanding the church’s message. And, finally, soldiers tended to distrust the church’s message because of its wide-spread reputation of being part and parcel of the ruling elite. The church failed, Cairns states, because it had a lack of the ‘Spirit of reality’, a ‘lack of love’, in the sense that there is too little fellowship in Christian congregations and too much class distinction, too little sympathy with the workers’ efforts for social reform which renders the church futile in the eyes of these men and, finally, a want of Christian unity. The church also suffered from
a ‘lack of life’ as it was out of touch with the modern life and its demands. Moreover, it not only failed to properly educate the men in religion, but also to prepare them for war itself. The material world of war led men to stop thinking and led the emphasis of their existence on their bodies: endurance, obedience, perseverance. This left little or no room for religion. Cairns emphasises that ‘If the Christian Churches acquiesce in the continuance of the present condition of things, they will be far more guilty in the sight of God than those who in ignorance deny Him.’

Schian also emphasises the distance between the soldiers and the chaplain. Like the British churches, the German Protestant Church was being criticised for being a Klassenkirche – acting only in the interest of the higher classes. Chaplains were seen as representatives of this high class church, and as part or even instruments of the Obrigkeit. This of course did not contribute to the credibility of the chaplains’ message among the soldiers. When the chaplain was seen as part of the establishment that had brought the war on, and that forced these men to suffer and die, how could the soldiers still trust him and the sincerity of his religious message? When was the chaplain speaking as a man of God, and when was he speaking as a servant of the establishment, acting as a moral booster, stimulating the men to fight on? ‘As soon as the critical men believed him (the chaplain) to be in the service of the political government “system”, all trust was gone.’ Schian states ‘that although the war has initially helped the church’s mission, it has increasingly become a powerful obstacle for the church’s work.’

This problem grew more pressing in the later war years when the men became war-weary. Had the chaplains won the trust of the men early in the war, they lost it again later when soldiers believed them supporting the lengthening of the war. Perhaps they were right, poses Schian, for one could certainly raise the question whether the chaplains had been caught in the expectations of 1914 for too long. This was not only a problem for the chaplains, but for the church as a whole. It left the church unable to react properly on the changing atmosphere among the German people. While a growing part of the German population grew to hate the war and those institutions that they felt were responsible for it, the Protestant Churches continued to support the state’s war effort.

It can be no surprise that the war loss and the subsequent harsh measures in the Versailles Treaty made this even more problematic. The church’s unbroken support for the state and the war led people to regard the church as responsible for the war, and the suffering it had caused. Schian quotes from a report based on the experiences of 41 local churches that ‘the majority, that is the young people, is possibly even more distanced and hostile than before. (...) Often, the bitterness against the government which had grown during the war is also aimed at the church.’

Looking back and forward

The authors of these three reports do not only look back. They also look forward: how can the problems they identified by countered in the future? Schian started his evaluation of the church’s war effort in 1916. At the time he finished his conclusions, the war had taken a drastic turn. Germany had lost the war, the German Empire, with which the Protestant church had been so closely related, was no more. Schian’s conclusions are clearly in the shadows of these important events. It is difficult for him to look forward. The future is insecure. Propositions of change in the organisation of the chaplaincy, such as a more independent organisation, stranded on the severe reduction of the German army. The position of the chaplains could hardly be top priority now. Schian states that the German chaplains had done the best they could under the extremely difficult circumstances they had to deal with. And, notwithstanding, the critique that Schian certainly had on the attitude of the churches towards war and the state, he claims that, ultimately,

*the position of the protestant churches during the war was – mainly – correct. They were convinced of the necessity and the consequent moral justification of the war. They believed the*
war a defensive one with Germany fighting for its life, which it really was up until its terrible conclusion. As such, they were right to bless the German weapons, to celebrate German victories, to call for perseverance, and to support the state's measures to make perseverance possible. They needed to act thus, while they are German protestant churches; and they could act thus, without forgetting their responsibilities as Christian churches.22

Looking forward is at the heart of Cairns work. To advise the churches on how to proceed after the war is the very agenda of his report. As essential points, Cairns claims the church should rethink its religious message, find better ways to answer the need of the men, counter their ignorance in religious issues and, finally, the church should take a more critical stance in society. It had to do this through example and precept, through fellowship, by practicing its own principles and teaching these to make them the principles of society as a whole. The church should have focused on social reform, on teaching society about the core of Christianity and on cooperation between denominations. After the experience of war, the world was beginning to look for light and guidance. A new opportunity for the churches: “The soil is crying out for the seed.”23

In the eyes of Grandmaison and Veuillot, chaplains and soldier-priests had successfully initiated a positive change by their work at the front. Through their exemplary Christian conduct, they had been able to show the nation the value of the Catholic creed and church. Moreover, war had revealed some structural flaws and shortcomings, but also illuminated hopeful signs of humanity, respect and, sometimes, of religiosity. The majority of the chaplains from Grandmaison and Veuillot’s report confirm that the war brought a certain improvement of the hearts, a superior quality in men. Organisational issues had to be solved, for instance in the shape of a senior chaplain as Grandmaison and Veuillot had seen in the British army. During the war, Grandmaison had personally lobbied for this. However, the French government had strong objections against such a senior chaplain as it feared a kind of spiritual police under the authority of Rome.24

Conclusion

Lessons should be learned from the church’s past failures. The German Protestant Churches emphasised that last July. All three sources discussed above are convinced of the relevance of this claim. But they differ on the nature of these lessons as well as on who should be the main pupil. Schian defines his lessons learned differently from his present-day colleagues. He recognises that the close relation between church and state had proven problematic. But on the whole, the church had not been wrong in its attitude towards war and the state. Practical matters should be changed in improving the chaplains’ organisation and in reaching more independence for the chaplains. But the justification of the church’s was support is not questioned.

For Grandmaison and Veuillot, the pupils are both state and nation. State should learn the need for spirituality, the value of Christian charity and moral and mostly the value of the French Catholic Church as a much needed spiritual guide of state and nation. The French people should have learned to abandon anticlerical prejudices and follow the soldier-priests’ example of charity and sacrifice. In Cairns’ report, the church is indeed the student. It should learn how to pass its message in a way that the ordinary man could understand and to find a position in society where it could represent the whole nation and not just a small part of it. The spiritual opportunity presented itself anew after the war.

Similar questions thus led to a variety of answers, which were highly influenced by the national contexts in which these answers were phrased. Comparing these sources shows just how important these national contexts and pre-war circumstances were for the way the churches evaluated their participation in war. We see that again in the fact that, although the British and French churches too called their nations to arms in August 1914, today only the German churches
feel the urge to make an official declaration distancing themselves from this part of their history. It is much more the shadow cast by subsequent events; the resistance against the Weimar republic, the rise of extreme nationalism and the Third Reich, that feeds the sense of shame and guilt which is now felt. But one thing did become clear to all three churches: war was never a straightforward spiritual opportunity. It was a learning school instead.

Endnotes

1 The nature of the phenomenon of war enthusiasm has been highly contested among historians.
6 Martin Schian, Die evangelische Kirche in der Heimat (E.S. Mittler & Sohn: Berlin 1925) 3.
7 Geoffroy de Grandmaison and François Veuillot, L’aumônerie militaire pendant la guerre 1914-1918 (Bloud et Gay: Paris 1923) 305.
8 Cairns, The Army and Religion, 74-75.
11 For instance, Innocent X issued a papal brief in which he advocated a canonical and hierarchical organisation of the chaplaincies. Though Spain adopted such an organisation, in France the brief was ignored. See Xavier Boniface, L’aumônerie militaire française (1914-1962) (ed. Du Cerf: Paris 2001) 28.
12 Grandmaison and Veuillot, L’Aumônerie Militaire, 216.
13 Ibidem, 228.
14 Martin Schian, Die Arbeit der evangelischen Kirche im Felde (E.S. Mittler & Sohn: Berlin 1921) 490.
15 Ibidem.
16 Grandmaison and Veuillot, L’Aumônerie Militaire, 233.
18 Schian, Die evangelische Kirche im Felde, 273.
19 Ibidem, 7.
20 Ibidem, 274.
21 Ibidem, 9.
22 Schian, Die evangelische Kirche in der Heimat, 27.
24 Grandmaison and Veuillot, L’Aumônerie Militaire, 256.
‘What was the reality for the average Frenchman under occupation and how did the Germans exert control of the population. As consent of the population lies at the heart of current COIN doctrine, what are the conditions for success.’

Group Captain D J TORIATI

Introduction

The clocks did not strike thirteen in Paris 1940. But moving the time zone forward to align with Berlin was just one of many forced changes that generated great stress across the population. France’s loss of control over time symbolized the reality of an occupation that would become almost Orwellian by 1944. Yet there remains a perception that the ‘average’ Frenchman was not overly burdened by occupation and that most simply wanted to live a normal life. German ideology and Pétain’s Vichy Government gave many the opportunity to do just that. But for others, a normal life was elusive. The privations of rationing, the enforced labour requirement of the Service Travail Obligatoire (STO) and the horrors of Jewish deportation and retributive massacres made life intolerable for many ordinary French people.

This paper explores three themes: the conditions for success in contemporary COIN; the reality for the average Frenchman under occupation; and the methods used by the Germans to control the population. It begins with an analysis of Campaign Authority using the CPA’s early failure to gain neither consent nor control in Iraq, to underline its relevance. The paper highlights the CPA’s lack of political understanding of Iraq, the purge of state institutions and heavy-handedness as factors behind this failure. In contrast to the dangerous and growing insurgency facing the CPA, Germany was to be more fortunate in France. It found in Maréchal Pétain a man who viewed collaboration as the only way to avoid penury for the State and a population that appeared to have settled for an ‘accommodation’, neither overtly collaborate or resistant in nature. But German control became heavier-handed from 1941 and their authority and legitimacy declined. Though a national uprising failed to materialise, acts of resistance became more prevalent as occupation continued. The resistance strengthened as German success on the battlefield faded, and efforts to support their war through French labour and resources became ever more desperate.

The consequences of resistance and collaboration meant that France could not avoid a national schism in 1944. It was salved only by the highly competent retribution management initiated by de Gaulle who drew on the spirit of La Résistance as a unifying catalyst. It is suggested that successful retribution management may be an enduring consideration for post-conflict management. Throughout the paper, the conditions for success in COIN are discussed. It is concluded that, understanding context, and tailoring the interventionist’s approach to a specific set of circumstances, are critical to achieving desired aims.

Iraq – the aftermath

The invasion of Iraq and the fall of Baghdad on 7 April 2003 occurred in just 18 days. What followed precipitated a vicious insurgency which the Coalition would struggle to contain over the next 8 years. Bremer, an American diplomat, was charged with administering the CPA. Having taken control, he enacted the most debated decision of his tenure; disbanding the apparatus of the State and the security forces in particular. His reasoning is discussed by Bob Woodward: ‘Bremer was reported as saying: “The thought is we don’t want the residuals of the old Army. We want a new and fresh army.”’ This ideology had been determined before the circumstances of Iraq were fully understood. Bremer would intimate that a broad consensus existed in the US for total regime
change stating that: ‘A memo (was sent) from Donald Rumsfeld on May 8, 2003, that said “the coalition will actively oppose Saddam Hussein’s old enforcers. We will make clear that the Coalition will eliminate the remnants of Saddam’s regime”.’

Regime change of this nature would have required a deep understanding of the underlying dynamics in the country, a secure, controlled environment and consent from the people for that control. UK doctrine is coherent with this view and considers consent a key condition for the success of COIN operations. It is one of four factors that determine Campaign Authority, namely: “legitimacy; the nature of the conduct of those exercising their mandate; the consent given by the population to those exercising their mandate; and the degree to which they meet the aspirations of the population”. Goldstein eloquently brings these themes to life:

“To succeed, [the occupiers] must win the hearts and minds of the occupied populations. They should also show the occupied population that they can provide security and a better future. If they don’t or can’t then they will face the ever increasing risk of insurgency”.7

The conditions required for success in Iraq simply did not exist. Wolfowitz and others had believed “ethnic tensions to be exaggerated” not only underestimating the seething resentment in the country but the role other actors, including Iran, would play in the ensuing conflict. The policy of imposed democracy invited retribution from the displaced Sunni minority. They quickly realised that there would be no way back into a majority, Shia-dominated elite, and no reason therefore to support the CPA. The approach lost the CPA its legitimacy and authority. The situation was exacerbated by those exercising their mandate like General Odierno, 4th Infantry Division, who was acquiring a reputation for heavy handedness:

“Instead of trying to secure the population, his commanders launched large-scale sweeps to roll up insurgents and Ba’athist leaders, fired artillery blindly to interdict insurgent activity, purposefully detained innocents to blackmail their insurgent relatives and levelled homes to deter people from supporting the insurgents.”

The issue was further compounded by a lack of relative mass. An Iraqi Army of 400000 soldiers and tens of thousands of Fedayeen, isolated, disempowered and unpaid, melted in to the cities, towns and desert, bitter at their treatment. Security became the responsibility of just four divisions of American soldiers in US controlled areas compared to 40000 German soldiers in Paris alone in 1940. A calamitous vacuum developed, filled by those disadvantaged by the CPA’s attempts to build a new state. The Coalition would not gain Campaign Authority until the US surge began in 2007 and a new approach addressed the lack of control and consent. The situation in post-invasion France was significantly different.

France 1940 - the aftermath

The Fall of France in 1940 ended with the signing of a debilitating and ultimately humiliating armistice in Rethondes on 22 June 1940. It was, with crushing symbolism, signed in the same railway carriage as the Armistice of 1918, a document that had helped condemn Germany to economic misery in the following decade. The German invasion, executed with ruthless efficiency, gave the Germans an unassailable air of invincibility in France. Any sense of defiance crumbled in the face of overwhelming power. Almost 2 million POWs were taken during the fighting with two thirds remaining in captivity until the end of the war, their families left headless for the duration of the occupation. An ‘Exode’ of millions of French people quickly followed, precipitated by fear of retribution, perhaps even hope of liberation or of a counter-attack that would shortly follow. But with a display of pragmatism and realism that would come to symbolise the average French experience, the refugees soon turned round and made their way home. They found towns and
cities that were largely intact, bureaucracies and institutions in place, a functioning infrastructure and a security force that was recognisably their own.

The first act of the Germans was to ‘zone’ the country through a process of establishing the economic, political and military significance of a region. Germany needed direct control over resources whilst managing the balance of the force required to meet this aim and thereby maximise ‘returns’ on their investment. To do this, they created an ‘Occupied Zone’ encompassing Paris and the North where the most significant agricultural, population, economic and political centres lay. Additionally, they took sole control of areas perceived to be of most military significance including the Channel and Atlantic coasts.

The second major zone was the Vichy administered ‘Free Zone’. By allowing the French to administer France from Vichy, the Germans were able to control large tracts of the country with the minimum of effort. And in Field Marshal Pétain, the ‘hero of Verdun’, the Germans were presented with an iconic figure to take control of Vichy France. In WWI, Pétain had witnessed devastation on an unimaginable scale and was unwilling to see suffering like it again. His personality and experiences permitted a relationship that the Germans saw as a unifying force, allowing him to preserve France and keep the population under control. Pétain made a conscious decision to collaborate believing it to be in France’s best interest to do so; he was the shield to de Gaulle’s sword. He saw himself as both realist and patriot, determined to avoid the worst privations for France following military defeat; it was a view that would set the tone for many of his countrymen.

German control

German victory in France was not followed by a systematic dismantling of the State. Instead, it could be argued that their intent was to collaborate with willing partners, gain consent and maintain legitimacy. Early events however appear to contradict these aims. The terms of the Armistice in particular were oppressive, forcing the French to pay for the costs of occupation and to support the wider German war effort: “it drained the country of goods”. Germany itself had been suffering under its own Armistice from the end of WWI and was in no mood to spare France the reparations it felt it deserved.

However the situation within France seemed to favour German success, with little need for further heavy handedness. The changes precipitated by occupation appeared to support the majority of Frenchmen with the country:

“Deeply divided politically, polarized between right and left. The right was eager to abolish liberal parliamentary democracy and replace it with a strong state based on obedience, deference, discipline and respect for democracy.”

Affirming the value of Sun Tzu’s maxim of ‘knowing the enemy’, Germany’s understanding of the French wider context allowed them to exploit the country with ruthless precision. And with no real desire nor need to replace French institutions with German facsimiles, they maintained them, gaining consent and control by doing so. The former SS intellectual and head of the government subsection of the Military Administration (MVW), Werner Best, had developed the concept of ‘supervisory administration’. He resisted initiatives that might antagonise Vichy, such as art seizure. This ideology is supported by the views of historian, Werner Rings, who identified a “general tendency and readiness of the inhabitants of occupied territories to compromise with the enemy for as long as humanly possible.” Supervising a compliant French state with a ‘light touch’ would, it was believed, support this tendency and reduce psychological stress on the occupied. Normalcy was stressed. French bureaucracy, and security services were “essentially untouched”. Culture was allowed to flourish and was actively encouraged, excepting of course activity that was
anti-Nazi. The German soldiers posted to France were often considered too old to fight on the eastern Front, stayed for longer and developed constructive relationships with the French. To reduce stress (and cost) further, the number of soldiers dropped quickly after the invasion. On the evidence, the thesis of 'light touch' as a method of population control begins to look more compelling.

However, problems remain with the argument. Even if the population was initially lightly supervised, the State itself was not. There is evidence that there was firm German direction for extant institutions. But there was also an expediency to this approach that would have been as attractive to the Germans as it was effective. The German bureaucracy lacked the administrative skill of the British whose experience in colonies like India had gained them centuries of valuable experience. The ability of the French to manage the more mundane aspects of day to day life with minimum German input was central to the economy of effort necessary to maintain the war effort on multiple fronts. On balance therefore, the German approach in France appears to have been 'light enough' to maintain consent and meet the aspirations of the population.

The reality under occupation

To many historians, France was not as damaged by the war as others and the population is pilloried for its approach; ‘France did not do so badly’ says Tony Judt with Smith opining that: “There is a disdain for the French’s eagerness to simply live through the occupation with little thought to outside events.” As ever, it was context that set the conditions for this approach. Pétain was desperate to avoid the destruction of WWI and the penury of ‘Polandization’. His vision prevented the devastation he feared most, but France’s economy and infrastructure would nevertheless, become increasingly damaged by the war and the reality for the Frenchman suffered accordingly. In particular, as the Allies closed in on liberation, France suffered terrible infrastructure loss: “During the fighting of 1944–1945, France lost the use of some 75 percent of its harbours and rail yards and half a million houses were damaged beyond repair”. This perhaps confirms Pétain’s worst fears that failure to forge an early collaboration would have seen France raised to the ground in 1940.

From the beginning, the occupation caused deep anxiety across society. A curfew was in place, the road signs were changed to German and Nazi flags flew where previously the Tricolour had been raised. People queued for everything from milk to lipstick, the numbers of births, deaths and marriages dropped whilst food prices rose and stocks fell; the occupation clearly burrowed deeply in to the French psyche. The shortage of food raised tensions across French society highlighting a common experience whilst exposing deep divisions. Everyone was subject to rationing but the extent of its impact differed across the country. Many city people began to resent country people more even than their occupiers, with food or lack of it an underlying cause. In the country there was more opportunity to grow food and less interfering supervision. Country dwellers salved their conscience and did their duty by sending food parcels to relatives and friends in the city. Ousby points out that “over 13.5 million ‘colis familiaux’ were sent in 1942”, making a key contribution to overall calorific intake for many people in France. To help ease the growing nutritional problem, the Germans divided the population into groups and allocated rations accordingly. But ultimately, the overall calorific intake dropped. The problem was compounded by the ‘Ravitaillement’ or requisition. German control over production and supply ensured desired goods were produced and sent to the Reich. Smith says that “the Germans were not interested in needlessly antagonizing the French, so food allotments were not intended to starve the local population”. But this lack of food, so central to French life, is symbolic of the oppressive nature of the occupation. It was probably the single most common reality across France.
However, working life for many people continued, with the Germans finding a largely supportive industrial base to help their war effort. By 1941, 7000 German firms were taking orders; in 1944 that number had doubled. Renault and others built machinery for the Wehrmacht making them vulnerable to accusations of collaboration. But manufacturers were employing people who would otherwise have had no work or been sent to Germany. Were these ‘working class’, average Frenchmen, collaborators in any pejorative sense? Where did ‘criminal’ collaboration start and ‘realism’ end during occupation? These questions only emphasise the ambiguities of life under occupation. As the workers and their companies looked to have had little choice, the conclusion can only be that it was realism that prevailed.

The reality deteriorates

Despite the privations and ambiguity of life under occupation, the French appeared determined to pursue as normal a life as possible: “French people just got on with it”. However, the ideology that allowed the Germans to control the population began to erode from 1941 on. The SS took responsibility for security from the French Army in 1942, and their policies undermined the MVW that had conspired to maintain Campaign Authority in the opening phases of the Occupation. The SS bombed Synagogues and appropriated art ensuring that relationships between them and the MVW: “would never be cordial in France”. Jews in the Occupied Zone were required to wear the Yellow Star from May of 1942 and, in August, the first Jews from camps in the Southern Zone were sent to Drancy en route for Auschwitz. From November 1942 the Germans occupied the Southern Zone compromising the Vichy Government’s attempts to appear independent. The effect of these events was compounded by a weakening perception of German invincibility as a military power. The invasion of Russia by Germany in 1941 changed political allegiances whilst the lifting of the Siege of Leningrad in 1943 and the Allied invasion of Sicily later that year eroded German authority further still. Hope that France had secured a lead role in a New European Order was fading rapidly and attitudes amongst the occupied were hardening against their masters.

In the small North West town of Évron, German flags were replaced with the Tricolour. Through small acts of subversion, such as the scrawling of graffiti and cable-cutting, the idea of resistance began to seep into the average Frenchman’s consciousness. This early resistance, of symbolic acts, of underground newspapers and leaflet drops, was more intellectual than physical in nature; “(the resistance) chose to make the first battle it fought a battle of language”.

The requirement for French citizens to work for the German industrial machine began to place a heavy toll on the population. Indeed the STO is cited by many as one of the most significant policies that led to the breakdown in French consent and the rise of a national resistance. Over 600000 men were required to go to work in Germany from 1942 onwards. Whilst the STO applied to a narrow age group of young men, the effect of their absence increased the pressure on families and communities, many of whom were suffering through the continued absence of POWs. Thousands more however, escaped the call-up and fled to isolated mountainous and rural areas. The image of the resistance most identifiable in popular culture is that of these ‘Marquisards’. However, they remained a disorganised, ill-disciplined faction rather than a large scale organised resistance at this point. At the time of the fall of France, perhaps the only group which was sufficiently organised and capable of setting up a clandestine resistance to the Germans was the Partie Communiste Français (PCF). However, this was also the time of the 1939 Nazi-Soviet non-aggression pact which “cast the PCF and Nazi Germany in to an uneasy alliance”. The Hitler-Stalin pact and the PCF had a loyalty to Moscow which precluded action against Stalin's German allies. Operation Barbarossa, the German invasion of Russia in 1941, opened the door for the PCF to change allegiance and they played an increasingly important role in resistance thereafter.
The average Frenchman appeared not to want the increased attention that came with violent resistance, underlining their wish to live as normal a life as possible. Hostage taking and executions by Germans in Nantes and Tulles\(^45\), and the horrific massacre in Oradour, vividly demonstrated the potential consequences of violent resistance. When 642 villagers were massacred for nothing more on their part than a suspicion by the Germans that they were harbouring resistors, the Germans proved that they could be vicious in their retribution, particularly if it was a response to the loss of German blood. Resistance carried too much risk for all concerned and without popular support it failed to generate significant mass or momentum. The reality for the average Frenchman was not, therefore, of overt resistance; it was increasingly clear that the main effort was to maintain normalcy.

The growing atmosphere of fear and mutual distrust corroded the seams that bound communities together and stymied the growth of resistance. At the political level too, the resistance lacked support. De Gaulle’s longer term aims saw him favour the spirit of a resistance rather than actual resistors: “Resistance was proof of La France éternelle, resistors were trouble.”\(^{46}\) Trouble in particular came from the PCF and a potential challenge to his power-base in a newly liberated France. Conversely, Vichy, openly in collaboration, feared German reprisals and the upsetting of the status quo. Over 500000 people were thought to be involved in the national resistance movement with up to 100000 known to have died.\(^{47}\) By today’s standards these numbers appear to be significant but they represented less than 2% of the population.\(^{48}\) Without state-level organisation or sponsorship, the resistance did not gain the traction that many might have expected. Neither the population in France, de Gaulle in London nor Pétain in Vichy gave the support necessary to foment a large-scale uprising.

**Retribution management**

‘Retribution management’ is central to generating the longer-term, post-conflict\(^{49}\) success of a campaign. De Gaulle and the Allies believed retribution to be an inevitable consequence of the approach of the French under occupation, and developed a strategic plan for post-war retribution management. In France, summary retribution against criminal collaborators was swift, taking place amongst the French around the D-Day landings.\(^{50}\) Indeed, most violent retribution had taken place by the time of VE day. Longer-term retribution management needed a Frenchman to ensure it was managed with the consent of the people. But de Gaulle was concerned right up to D-Day that the allies would seek to take power rather than him:

> “He feared that Eisenhower’s address and proclamation to the French, which made no mention of General de Gaulle and the French Committee of National Liberation, foreshadowed “a sort of taking over of power in France by the Allies’ military command.”\(^{51}\)

It is likely that with a ‘provisional authority’ in control, the reality for the Frenchman from 1945 would have been far worse than it became. De Gaulle had cultivated an image of legitimacy and broadcast a unifying message that drew on the concept of a national spirit of resistance. This approach, coupled with an overall effort to ‘let bygones be bygones’\(^{52}\), held sway for 20 years following liberation. It was not until commentaries including Marcel Ophüls’ documentary ‘The Sorrow and the Pity’\(^{53}\), were published and subsequently banned in 1969 that the reality began to play out on a wider stage. By then, France was strong enough to begin an introspective examination; any earlier and the consequences may have been calamitous.

**Avoiding ‘templates’**

The different strategic aims and ideologies in Iraq and France limit the extent to which meaningful comparison of the Campaigns can be made. Moreover, the reality for the French and Iraqis
changed markedly as the respective occupations wore on. However, it is possible to conclude that the early strategy adopted by Germany in France may have mitigated some of the manifest risk in Iraq. The opportunities presented to them by Pétain and others allowed the Germans to control the population by gaining Campaign Authority, including the establishment of a legitimate ‘French-faced’ administration supervised by the MVW. But each campaign is characterised by its unique context. It may not be possible, for example, to sustain a manufacturing base that allows an economy to endure, much less maintain an education and cultural foundation. And whilst establishing partnerships with state institutions may reduce the overall management burden, collaboration generates its own longer-term consequences. Partnering is problematic. ‘Proxies’, such as rebel groupings, may have political allegiances that are fickle, expedient or highly unstable. Aligning with a corrupt or unrepresentative government is also fraught with difficulty. In France, the Germans were presented with a set of circumstances that allowed them to effectively govern a compliant population. Today, in Iraq and elsewhere, significant sections of the population have determined that there is little to be gained through collaboration with Governments that lack legitimacy. The conclusion may be that, rather than dogmatically adopting a pre-determined ideology or template, future operations will need to be carefully tailored to a specific set of circumstances.

Conclusion

Like many contemporary historians, Clark’s view of the French during occupation is disparaging. Contrasting Russia’s approach to potential occupation with that of the French he says: “It was the antithesis of the apathy and resignation that lay behind the French defeat in 1940. Then a people sacrificed their country and institutions for their own personal safety.”\(^{54}\) It is, however, a simplistic view of a complex reality. French society came under great pressure during the occupation. Many were deprived of the main breadwinner through the STO or the absence of POWs. Most went hungry and lived under the oppressive regime of a foreign occupier, notwithstanding the existence of a French-fronted administration in Vichy. But whilst location, gender, race and age made a difference to the reality, on balance, Germany’s touch was ‘light enough’ to ensure in large part, the ongoing consent of the population.

By 1941 Campaign Authority was breaking down. The STO compounded the effects of the Armistice that blighted French life. Heavy-handedness, predominantly by the SS, lost the Germans legitimacy and authority. There became a growing realisation on the part of the French that the aura of invincibility of the Germans was also fading. With the erosion of their authority, and a change in the political landscape, a resistance took hold. It started as an intellectual exercise but grew in size and stature becoming visible enough to allow de Gaulle to use its spirit as a galvanising force during France’s post-war, retribution management phase. Strategic messaging enabled France to unite and recover enough self-confidence to quickly put the past and their reality behind them. This suggests that retribution management relies upon a strong message delivered by a leader with legitimacy on his side; it may well be one of the more important lessons to be learned for management of today’s campaigns.

Endnotes

2. United Kingdom. Development, Concepts and Doctrine Centre. British Defence Doctrine. Joint Doctrine Publication 0-01. 4th Ed. (Shrivenham: DCDC, 2011), 1-21. Campaign Authority is defined as: “The authority established by international forces, agencies and organisations within a given situation in support of (or in place of) an accepted (or ineffective, even absent) indigenous government or organisation”.
8 DCDC, JDP 0-01, 1-21.
12 Ibid.
13 Ricks, Fiasco, 117.
14 Ousby, Occupation, 111.
15 Laub, After the Fall, 2.
16 Goldstein, Military Occupations, 47.
17 Ibid.
18 Militärverwaltung in Frankreich.
19 Vinen, the Unfree French, 109.
20 Laub, After the Fall, 76.
22 Goldstein, Military Operations, 48.
23 Meredith Smith: "the civilian experience in occupied France, 1940-1944" (2010), History Honors Papers. Paper 6. http://digitalcommons.conncoll.edu/histhp/6, citing Gildea, Marianne in Chains, 255: “they (the Police) found themselves simply doing the dirty work of the SS under the guise of independent action”
26 Smith, Occupied France, 75.
27 Robert O. Paxton, Vichy France: Old Guard and New Order, 1940-1944 (New York: Alfred A. Knopf, 1972) 359 cited by Smith: “Polandization refers to the extreme sufferings of the Poles under Nazi occupation. It was often used as a reference to contrast the experiences of occupied nations”.
28 Judt. Retribution, 294.
30 Vinen, the Unfree French, 231.
31 Ibid.
32 Ibid., 222.
33 Ibid., 231.
34 Smith, Occupied France, 14.
35 Ibid., 15.
37 Smith, Occupied France, 17-18.
38 Davies, France, 13.
39 Laub, After the Fall, 133.
40 Ousby, Occupation, 328.
42 Ousby, Occupation, 219.
43 Vinen, the Unfree French, 5.
44 Laub, After the Fall, 113.
45 The death of the Feldkommandant in Nantes in October 1941 led to the execution of 150 Frenchmen ordered by Hitler in retribution. 200 civilian men died in Tulle following attacks on Nazis in the city. Ousby, Cobb et al.
46 Ousby, Occupation, 300.
48 Ibid., 2.
49 Judt, Retribution, 301.
51 Laub, After the Fall, 2.
52 Marcel Ophüls, The Sorrow and the Pity. DVD. 1969 Productions Television Recontre S.A.
What factors explain the Italian participation in multinational military operations over the last decade?

Major Dario Paduano Italian Army

Introduction

Since the end of the Second World War Italy has participated in 124 military operations, of which, 33 are ongoing in 25 countries.¹ The first Italian post-war military operation was in Somalia in 1950. Based on a UN mandate, it ended in 1960 with Somali independence. Over the last decade Italy has sharply increased its participation in international military operations. Until the late 1980s the main characteristic of these interventions was their small military contingents. The only exception to this was the peacekeeping operations in Lebanon between 1982 and 1984. Italy's involvement has grown exponentially from four deployments in the early 1980s, to nine in the latter half of the decade. By the 1990s, Italy was participating in more than 20 per year, reaching 30 participations by 1999. Since 2000, the number of Italian overseas operations has remained steady at about 30 per year.² Italian deployments have not only grown in number but also in size. The 1990s saw Italian multinational operations increasing, starting with Operation Desert Storm, following the Persian Gulf crisis (1990-1991), and then contributing to all operations in the Balkans (especially between 1995 and 1999). Therefore, since her first military operation in Somalia in 1950, Italy has increased the use of her armed forces abroad.

This is considered an extremely important subject, for the following reasons. First, both the number of operations and number of deployed troops has increased significantly. In the 1980s this was just a few hundred personnel, by the 1990s it was a few thousand, and by the first decade of the twenty-first century it had risen to an average of 8000 deployed contemporary.³ Therefore, the presence of military personnel overseen has become significant, recurrent and sometimes a contentious issue in both officials statements and across the media. Second, as troop numbers have increased and the operations have grown in complexity, the cost of the operations to the Italian taxpayer has also increased from around €461 million in the 1980s to €2.087 billion in the 1990s (of which 25% was spent in 1999 alone) to €3.91 billion in the 2000s.⁴ Meanwhile the overall Italian defence budget (that is in addition to the cited costs) has reduced from an average of 1.5% of Gross Domestic Product (GDP) in the 1990s to an average of 1.3% of GDP in the 2000s.⁵ Third, with increased deployments Italian soldiers have been exposed to greater risk, and the number of soldiers killed on operations has also increased from just 1 in the 1980s, to 35 in the 1990s, to 95 fatalities between 2000 and 2013.⁶ These figures, in combination with the number of non-battle injuries,⁷ have increased awareness among the population about Italian participation in military operations. Finally, the employment of the Italian armed forces abroad, when there is no direct threat to the State, has raised a constitutional debate. Article 11 of the Italian Constitution states that “Italy repudiates war as an instrument offending the freedom of other peoples and as a means for settling international disputes...” this has always created, within coalition governments, an issue about the legal framework within which any Italian participation should have been justified.⁸ Italian governments, facing potential attacks from opposition parties and some sectors of the electorate, have tried to avoid any constitutional debate about the use of the armed forces. They have preferred to emphasise that participation of the Italian armed forces has taken place under the umbrella of either an appropriate United Nations Security Council Resolution (UNSCR), or in concert with close Allies. Moreover, officials have presented all military interventions as a national contribution to ‘peace operations’, providing humanitarian assistance, or to protect human rights.⁹ Understandably, the cost of military operations in terms of both ‘blood and treasure’, and the difficult legal framework to justify such actions have led to heated debates across Italian society.

Yet this important topic has failed to energise attempts in academia or the media to explain the policymakers’ decisions and, therefore, this paper will contribute to the literature on this subject. The
paper will look to explain the motivations of Italian decision makers when deploying Italian military forces in support of multinational operations. The causes of key military interventions will be analysed through the prism of International Relations (IR) theory, this theory was first expounded after the end of the First World War which signalled IR theory as a discipline. However, since 1918 numerous academics have developed IR theories to try to analyse and explain key decisions, including the causes of military intervention. Nevertheless, as William Lahneman points out, “cultural norms, political dynamics and economic backdrop make each cases of Military Intervention unique”, but this does not compromise the value of IR theories as explanatory tools for this phenomenon. Realism and Liberalism are the two mainstream IR theories and they will be used to explain states’ behaviour, because they present a different vision in explaining what can affect policymakers. They look at world politics through different ‘lenses’ and assume as factors of influence two different categories of causes: internal and external. The Realist theory proposes that military intervention is a reaction to external inputs, while the Liberal theory explains a state’s actions as a reaction to domestic pressures. The realism ‘brand’ to which this paper will refer is Stephen Walt’s neo-realist theory. Walt explains states actions as a reaction to threats which he defined in terms of “aggregate power, proximity, offensive capability and perceived intentions”. While the Liberal paradigm that will be utilised is crafted by Andrew Moravcsik, who explains a state’s conduct through international politics acting as a consequence of the relationship between the State and its domestic social context. Using IR ‘lenses’ may explain military intervention, however, Italy has always participated alongside its Allies, both as a member of International Organizations, most notably the UN, EU and NATO, but also as a member of the so called “coalition of the willing”. Thus, Walt’s account as to why states join alliances (a balancing versus bandwagoning approach) will complete the theoretical explanation of those factors that can influence policy makers, when making decisions about participating in multinational military operations.

Using these two theories as a tool to explain a state’s actions this paper will answer the following question: should the decision by Italian policymakers to participate in multinational military operations be understood as a reaction to international pressures or as a result of domestic drivers? To answer such a question this paper will consider both the internal and external factors of influence over Italian statesmen. Internal factors of influence are those generated by domestic social groups, public opinion, lobbies, powerful economic actors, coalition parties and opposing parties. Conversely, external factors of influence are those standing from influential state actors, allies, international organisations and threats, as defined by Walt. The findings will show that, contrary to what one would expect, external pressures prevailed at the detriment of economic interests and social values, raising public discontent and nurturing parliamentarian debate within coalition governments. The evidences will make clear that Italy complied with external pressures mainly to increase her relative power with important allies and enhance her influence and prestige within multinational organisations. Thus, the IR theory that better explains Italian behaviour and choices, related to the military operations that will be considered, is the realist theory.

Two case studies will be used to show why succeeding Italian Governments have been open to increasing the use of their armed forces as an instrument of foreign policy. Each case study will analyse the effect of both internal and external factors on policymakers. The findings underpin the argument that is the prevalence of external determinants in using armed forces as an instrument of foreign policy. The first case study will look at Italy’s participation in Afghanistan (which commenced in 2001 and is still ongoing). This has been selected because in terms of the resources allocated, personnel deployed, soldiers killed in action, and constitutional debate it has been far and away the most hotly debated since the end of the Second World War. The second case study will look at operations in Libya in 2011. Similar to the participation in Afghanistan, the second case has been selected because Italy’s participation created domestic turmoil. Public opposition was determined by kinetic targeting, air shelling and missile strikes used by a coalition first and then by NATO. Also, in this case, the constitutional legitimacy of this operation was hotly debated.
So this paper, using both primary and secondary sources, mainstream theories and scholarly insight, will assess political attitudes, preferences and choices (section I). It will go on to examine two specific multinational military operations to identify what determined Italy’s participation in each case (sections II and III) before extrapolating the conclusions that could interpret Italy’s willingness to join other multinational military operations. The conclusions of this paper lead to broader considerations about how effective external pressure can be when policymakers do not perceive the necessity, or ambition, to increase a state’s relative power. The timeframe examined will be the beginning of twenty first century from 2000 to 2013, as this is the period when this topic assumed the aforementioned relevance.

Military intervention: IR mainstream theories and alliances hypothesis

This section will start by outlining the key facets of the two mainstream IR theories (realism and liberalism) and how this can be used to explain a sovereign state behaviour. Specifically, it will seek to identify both the internal and external factors that can influence a state’s behaviour. A theory per se is not just a system that provides a set of assumption and hypothesis, but it is a ‘device’ that enables the identification of which elements are relevant and which are not. Through an understanding of what really matters for decision makers it is possible to categorize what can affect their decisions. The IR mainstream theories used in this section will not describe the causes of war, but will explain what factors can affect a statesman’s decision to apply military force, within a multilateral context, as an instrument of foreign policy. Overall this section will support core understanding of the issues that surrounded Italian participation in the proposed case studies.

The following concepts are articulated by Andrew Moravcsik and are slightly different from the classic liberal core beliefs that argue that peace is enabled by democracy, global economic ties and international organisations, formulated by the founders of liberalism, Adam Smith and Immanuel Kant, and supported by thinkers such as Robert Keohane and John Ikenberry. Moravcik’s main tenet is that what impacts on a state’s behaviour “in world politics” is the relationship between the domestic social context and state. A government’s preferences and, therefore, its strategic choices, are influenced by ideas, institutions and interests that create the national context. But these ideas and interests are not universal even within the same state, because of the different conditions within which people live. Actions are determined by the desire to benefit, therefore, the bigger the perceived benefit is the stronger the incentive to act in a certain way. Thus, a state can be considered to be a subset of its society and a government will act, in accordance with the society interest. Nevertheless, a state’s decisions are not necessarily the result of the entire society, rather within a society the ideas and interests of the strongest (most influential) group should prevail. Consequently, not all the groups have the same influence over the decision of a state, because each government is inevitably the expression of a group more than another. Social context will shape groups’ perceptions and as a consequence will shape their sensitivity about security or, for instance, their views on sovereignty. In this respect “states’ actions are determined by groups’ preferences, and the latter are determined by a set of perceived risks and opportunities”. Moravcsik propose three major variants of this theory: Ideational, Commercial and Republican Liberal theories. All of them explain state behaviour as a consequence of domestic factors.

Ideational Liberalism describes states’ preferences as an outcome of social values and group identities. The decision to cooperate or wage war on the international stage is shaped by these values and identities. Social identity set preferences includes collective goods, such as legitimate political institutions, national unity or socio-economic regulations. This variant of the Liberal theory identifies security, sovereignty and territorial integrity as purposes (i.e. not ends) to realise the preferences defined by social groups. For that reason, foreign policy will be crafted in accordance with the social ideas of groups related to the concepts of security, sovereignty and legitimate borders.
Commercial Liberalism is the second variant of the liberal theory. Individual, but also collective, conduct of states is based on market incentives and, therefore, economic transactions. Moravcsik argues that liberalism is a theory and not an ideology. Therefore, he disagrees with Richard Cobden's first image of liberalism whereby states interdependence, free trade, prosperity and individual liberty are determinants of peace. Commercial Liberalism rejects the idea that the economic incentive enables collaboration among states and, consequently, it does not necessarily create conditions for peace. Any change in the economic structure at national or global level can alter the benefit of the economic transaction. When this alteration adversely affects or improves the income of powerful private actors, they create pressure on governments to hamper or facilitate those transactions through political, economic or security initiatives. In other words, based upon economic opportunity or risks private actors can influence government foreign economic and security policy. The bigger the economic interests of powerful private actors, the bigger will be their incentive to influence government decisions and, therefore, state foreign policy. This concept is most apparent in the manner in which powerful oil and gas companies exert significant influence over statesmen. The latter can answer to these pressures threatening economic sanctions (i.e. embargos) towards economic competitors or with military intervention in a rich oil or gas country. In this respect military intervention is never justified in terms of pure economic advantages, but it is always hidden behind the necessity to preserve the so called ‘national interest’.

The last variant of liberal theory is Republican Liberalism. In this Republican Liberalism the main determinant of a Government’s policy is the domestic group that it represents. So, the policy will be in favour of the coalition government or powerful domestic groups. As a consequence, few will take an economic gain from others, without giving something in return to those that have contributed to create the benefit, most notably tax payers. Also, this proposition illustrates the economic support or tax concessions that a lobby (national industrial groups including national defense industry) can obtain to the detriment of the collective (again tax payers). Therefore, liberalism, in its three variants, makes it clear that decision makers and their foreign policy choices are influenced by internal issues. On the other hand, realism provides a contrary perspective.

Realism argues the occurrence of external matters as main influential factors. The realist founder’s (Hans Morgenthau and Kenneth Waltz) and thinker’s (Robert Pape and John Mearsheimer) core belief is that self-interested states are in competition with each other for power and security. There are different types of realism: Classical, Structural and Neoclassical realism, but all of them are focused on the concept of power and its distribution in an anarchic system to explain a state’s activities. Therefore, realist founders and thinkers do not really explain military intervention in humanitarian, peace support and counterinsurgency operations as an instrument of foreign policy.

Rather it is Stephen Krasner, who starts from an adapted neo-realist perspective, who best describes the important aspects of intervention. Krasner does not reject the realist main tenet: the necessity to maintain or recreate the balance of power on the international stage, and that states take action to safeguard national security. The intervention is generated by insecurity, created by the anarchic system. To mitigate the risk associated with insecurity and uncertainty, states collaborate with each other, especially when security concerns coincide, and such collaboration will improve their own security. Therefore, when states share the same security concerns they are willing to collaborate to improve their own security. The necessity to improve state security is so strong that it overrides the concept of sovereignty and, because national security is paramount, realists do not accept compromise or set limits on this issue. So, military intervention can be motivated by instability in a neighbouring country if this affects security in your own state. This can be used to explain a military intervention in an adjacent country, where there is a minority or part of the population that is harassed by a tyrant, or threatened for religious or ethical reasons. These kinds of situations can generate massive migration or refugee concentration at the border and, therefore, directly affect the security and stability of a bordering country. On the other hand, intervention can be determined by the requirement to increase
national prestige, which is typical of western interventions during the colonial period. It is difficult to decipher if decision makers are motivated more by the necessity to assure national security rather than national prestige, or even the need to maintain the international balance of power. The main critique of this realist approach is that military intervention is often undertaken for territorial gains, or to exploit the resources available in the target country. Therefore, the potential threat to national security and the necessity to maintain or re-establish a balance of power are window dressing to mask the real motivation.

A further refinement to the balance of power theory is provided by another neo-realist, Stephen Walt. He explains why states opt for alliances and the implications of such decisions. He argues that the necessity for states is to balance against an external threat. Such a threat may come from a powerful neighbour or an unstable third world country, or from a country that, because of its beliefs (ideology or faith), offer safe haven to groups which threaten other countries. In short, states join alliances to face regional external threats. This is the “balancing” principle argued by Walt, where states team up to tackle a common threat in order to mitigate a lack of national resources. Examples are NATO, the EU and the Anglo-French Defence and Security Treaty. The alliance with a dominant power, to improve the security condition, implies the recognition of his benevolence but also the acceptance of his influence. The influence of the strongest ally over the others explains why he accepts an offer of help. This influence grows in relation to the size of the gap of resources between a stronger ally with its weaker partner(s). Small states may have little influence within an alliance because their contribution adds little value (i.e. power) to it. An example is the unequal but mutually beneficial partnership that most European countries enjoy with the US. Ideology also provides a reason for states to ally with regional or global superpowers, because common values minimise the risk of smaller states being dominated by the superpower member. Walt’s second approach to explain alliances is “bandwagoning”. In this case the alignment is “with the source of danger”; the most obvious example is the Warsaw Pact where Eastern European states joined the Pact so as to not be invaded by the USSR. It should now be clear as to the range of factors that can influence decision makers and, consequently, a state’s conduct in terms of foreign policy, including the decision to participate in multinational military operations.

Italian participation in the multinational military intervention in Afghanistan

This case study will explain what led to the Italian military commitment in Afghanistan, from 2001 to 2013, and the Italian participation in both the US-led Operation Enduring Freedom (OEF) and the NATO-led International Security Assistance Force (ISAF). It will focus on three subsections. The first will give the context within which the Italian participation was included. The second will state the domestic issues considered by political elites. The third will identify the outward pressures, arguing their dominance over officials’ choices. It would be restrictive, considering a time frame of more than a decade, to only make clear what determined the Italian participation. Hence, this study will also consider other important decisions taken by the national authorities. Those decisions are related to the build-up of forces, the deployment of combat assets, the assumption of new responsibilities and, finally, the removal of national caveats. The research questions addressed by this section are: Why did Italy decide to commit such a quantity and quality of resources in Afghanistan? What influenced successive governments, with different political orientations? Were social values, economic concerns or coalition parties’ interests the dominant factors, or were decisions taken to comply with allies and international organisation requests? Through the use of primary sources this submission will provide evidence of the positive effect achieved by powerful allies, through their influence over Italian choices. In doing so, this section will validate the paper’s core argument: statesmen decided to participate as a reaction to external pressures.
Context

Following the terrorist attack against the Twin Towers in New York on 11 September 2001 a Multinational Coalition, led by the US, launched OEF, a campaign against international terrorism and based upon UNSC Resolutions (UNSCR) 1368, 1373 and 1386. OEF was aimed to unravel and destroy the ‘Al Qaeda’ terrorist organisation, prohibit access to and use by terrorist groups of weapons of mass destruction (WMD), and deter certain countries from continuing to support, directly or indirectly, international terrorism. Seventy countries offered assistance to OEF, 27 of which offered force packages to be used in the conduct of the military operations.

For Italy, the operation began 18 November 2001 by sending a fleet called the ‘Italian Naval Command group’, composed of the aircraft carrier ‘Garibaldi’ (carrying a Squadron of Harrier STOVL attack aircraft), the frigates ‘Sapphire’ and ‘Maestrale’, the patrol craft ‘Aviere’ and the support ship ‘Etna’, (later replaced on 18 March 2002 by the destroyer ’Durand de la Penne').

During the summer of 2002, the US asked Italy to make available an infantry Tactical Group to be integrated into the Coalition. On 2 October 2002 the Italian Parliament authorized the participation, starting from 15 March 2003 with a mandate of six months, of a military contingent of 1,000 soldiers called Task Force (TF) ‘Nibbio’. The mission assigned to TF Nibbio was of territorial control, neutralization of pockets of terrorists, and possible logistical bases for recruitment in the western province of Paktia. In the law decree that the government sent to parliament it explicitly asked for the approval of the Military Penal Code of War. TF Nibbio ended its operation on 15 September 2003. On 3 December 2006 the Italian naval units handed over their responsibilities to the US Navy and returned home December 17, marking the end of Italian participation in OEF.

On 20 December 2001 UNSCR 1386 authorised the deployment of ISAF into the city of Kabul and its surrounding areas. ISAF was a stabilisation and security mission, very different and clearly separate from OEF. Italy was one of the 7 countries, all NATO members, who contributed to this new task with a contingent of 250 soldiers. When NATO launched a new phase of expansion outside the capital in 2005, the Italian government agreed to assume lead responsibility for Regional Command West (RC-West) in Herat in Western Afghanistan. The Italian government imposed strict geographical caveats on its troops, to ensure they could only undertake operations in Herat and Kabul. Any request to deploy Italian troops outside these areas could be authorised by Rome within 72 hours, so that parliament could be consulted. Assuming the lead of RC-West did not imply only Command and Control (C2) responsibility but also more terrain to control, therefore, more maneuver units to deploy and the responsibility to teach, coach and mentor officers of the Afghan National Police (ANP), the Afghan National Army (ANA), and the Afghan National Border Police (ANBP) operating within RC-West battle space. The terrorist groups in the area reacted violently to the Italian presence as it restricted their freedom of movement. To face this threat the Italian authorities authorised the deployment of armoured fighting vehicles, attack helicopters, Unmanned Aerial Vehicles (UAV) and ground attack aircraft. From a contingent of a few hundred men and women in 2003 Italy surged to 2,700 in 2008, then 3,150 in 2010 and finally almost 4,000 in 2012. The next subsection will examine the reasons why Italy decided to deploy her troops to Afghanistan and whether there were any security or economic interests to safeguard.

Internal factors

The most recent Italian ‘Defence White Book’ was endorsed by the then Minister of Defence, Antonio Martino, in 2001. This document was focused on the restructuring of the armed forces, it did not identify any national security priorities and, therefore, did not address any security issue. The document that has addressed the priorities, risks and threats to Italy, and remains in force today, is the ‘Strategic Concept’ endorsed by the Chief of the Defence Staff in 2004. In this document only one
threat is identified, international terrorism, and yet the document did not consider the likelihood of a terrorist attack as significant. Moreover, Afghanistan was not identified as a country where Italy has or had any strategic interests. Indeed one of the main criticisms of Italy's use of national resources and, in particular, the armed forces in Afghanistan is exactly that, there does not appear to be any commercial or economic interests for the country. So there is clearly no national strategic interest, no direct military threat, nor any vital economic or commercial interests for Italy in Afghanistan. Thus, these aspects will no longer be considered as factors likely to have influenced Italian decision makers. Moreover there was, and is, public discontent in Italy for the military presence in Afghanistan which will be explained below.

The costs involved in deploying and keeping Italian troops and equipment in Afghanistan has soared from €36 million in 2001, through €356 million in 2008, up to €828 million in 2012 creating inevitably public dissatisfaction, particularly during the recent financial crisis. However, the major reason for public opposition to the mission in Afghanistan was a cultural reluctance to use military power and also the significant number of people killed in action.

Italian public opinion has, for many years, been opposed to the use of the armed forces outside the country for many reasons. First, participation in the Second World War and the direct experience of foreign occupation raised awareness about the enormous suffering that can be caused by the use of military force. Second, Roman Catholic culture, not least the presence of the Vatican City within Rome, has always had influence over the government. Third, in some sectors of society there remains a deeply held communist belief dating back to the Second World War and the post war period, which is consistently opposed to military operations. Finally, Italy’s controversial colonial past has reinforced a national aversion to the idea of deploying military forces in another sovereign state. These facts have contributed to the creation of a pacifist culture that compromises public support for some government decisions. Moreover, the combative nature of OEF and the ISAF deployments have fuelled the constitutional debate over interpretation of Article 11 of the Italian Constitution. Finally, political grandstanding and a deep sensitivity created by casualties has forced the Ministry of Defence to report publicly to parliament about the circumstances in which soldiers have been killed, the specific details reported in the press have added to public concerns about the deployment.

In summary, internal factors were not significant in the decisions taken to deploy Italian forces to Afghanistan for either OEF or ISAF. No economic issues or direct security threats needed to be addressed. Furthermore, considering popular opposition, coalition governments from across the political spectrum have risked losing the consent of the people by confirming, or increasing the Italian military presence to OEF and ISAF. It would certainly have been a vote winner for political parties to say that they would reduce or even withdraw forces if they got into power yet Italian governments have consistently adopted the opposite course of action. None of the Moravcsik’s liberalism variants explain Italian decisions. Even though public opinion is opposed to the use of armed forces abroad the Italian government was not influenced by these social values and identity. As a theory, Ideational Liberalism is incapable of clarifying official decisions. Furthermore, the financial and human resources allocated to OEF and ISAF did not create beneficial economic conditions, as the powerful economic institutions would have liked, nor have any economic gains been identified. For that reason even Commercial and Republican Liberalism fail to elucidate why Italy participated in Afghanistan and why Italy was so keen to assume such an important leadership role. Thus, Italian decisions were not the result of domestic drivers, but were they a reaction to external pressures?

**External factors**

Italy has participated in multinational military operations in Afghanistan for more than a decade. Notwithstanding this has been Italy’s most significant presence in a foreign country, during the Italian Republic, it is hard to justify this action in terms of strategic doctrine due to the lack of any such
document, issued at the highest political level. Such a document would help to understand the criteria for the employment of the armed forces, and why Italian armed forces are deployed in state “A” rather than in the country “Z”. In this strategic vacuum it is necessary to look at Italian foreign policy and, in particular, policy regarding the use of the armed forces as an instrument of foreign policy. Then it is important to look at any evidence that key allies, in particular the US, held sway over Italian policymakers. Finally, it will explain how Italian decisions were made to ensure Italy retained its relative power within international and regional institutions and hence avoided the risk of marginalisation.

The thread of Ariadne that runs throughout Italian foreign policy is the bipartisan consensus over the necessity to maintain strong relationships with both the US and Europe. This bipartisan consensus agrees the necessity to rebalance power within alliances, and on the need to improve national prestige on the international stage. Italy’s participation in multinational military operations is seen as one way in which Italy can retain and improve this influence. Italian government parties of the centre-right and centre-left have, since the end of the Cold War, tried to balance Atlantic and Euro bonds. The thing that really changed between the two opposing political camps has been the censure over the fighting of Italian soldiers that was loosened by the center-right government in 2008. For the first time journalists were allowed to go to Afghanistan and be embedded in military units, experiencing the routine in a Forward Operating Base (FOB) in Bala Baluk and Gulistan deserts or in the Bala Mourgab mountains within the RC-West area of responsibility. The aim of this policy was to capitalise on Italian efforts in Afghanistan and give internationally visibility to Italian efforts in support of ISAF. Yet this does not answer why opposing political parties can agree on military deployments. The answer lies in Italy’s need and dependence on alliances. Italy’s economy is based on trade and energy supply, it needs to secure commercial routes and stabilise geographic areas where this national business takes place. Italy is unable to guarantee the necessary security and stability with its own strengths and relies on international organisations, such as NATO, and especially on the close ties with the US. Additionally, since the fall of the Berlin wall Italian politicians have decided to play a more active role on the international stage, within NATO, the EU and other multinational organisations (such as the UN and the G8), to ensure that Italy is not marginalised. The subsequent paragraph will briefly give reason for the Italian’s concern of marginalization by introducing events that created such an anxiety.

Although Italy has been the World’s 7th largest economy and the 5th largest contributor to NATO it has felt that it has been marginalised by its allies. It is not the purpose of this paper to explain this phenomenon but it is necessary to give some examples which support this view. In 1994 Italy was not included by her allies in the contact group to deal with the Yugoslavian crisis, an area that is just a few nautical miles from Italy across the Adriatic Sea. This in spite of the fact that NATO had targeted the Serbian regime and launched these attacks from Italian air bases through airspace deconflicted by Italian air traffic control. In1996 when the then French President, Jacques Chirac proposed to the US President, Bill Clinton a redefinition of the NATO Command Structured (NCS), changes which would directly affect NCS Headquarters in the peninsula, Italy was excluded from the debate. When NATO redefined its relationship with Russia again Italy was excluded, because the contact group, based upon a proposal by Chirac, was just France, Germany, Britain and the US. In 1997 the collapse of the Albanian political establishment raised security concerns for Italy but NATO allies, most notably the UK, adamantly rejected any calls for NATO involvement in Albania while backing for the IFOR, a NATO-led operation in Bosnia. Finally, when the allies discussed NATO enlargement in 1997 the Italian proposal to include Slovenia and Romania took second fiddle to the German proposal to include Poland, Hungary and Czech Republic. These are just some of the events that reinforced the views of Italian political authorities about the necessity to increase Italian influence and standing within existing alliances and groupings to ensure the reversal of this trend, and to ensure Italy gets the appropriate level of respect for its actions. How to achieve this goal was clear to every statesman, Italy must use her armed forces within multinational organisations.
Examples of this include President Carlo Azzeglio Ciampi, the President of the Italian Republic, who, on 21 December 2004 during a ceremony to exchange wishes for Christmas and New Year with the Courts of the Republic, expressed the Italian nation’s “feeling of deep gratitude for the way the Armed Forces hold high prestige for the name of Italy in the World”. 82 Ciampi was not an isolated voice, his successor to the Quirinale palace, President Giorgio Napolitano, on 21 December 2011 during a video-conference to exchange wishes for Christmas and New Year with Italian contingents in Afghanistan, Libya, Lebanon, Kosovo stated “…today you and others before you, have given a great contribution to a renewed prestige and credibility of Italy”. 83 The writer Chiara Alleva, in a series of interview conducted from May to November 2007 with Ministers and the Undersecretary of Foreign Affairs and Defence, enquired about the Italian use of armed forces. Members of parliament of both the centre-right and centre-left highlighted the boost given by the armed forces, employed in missions abroad, to national credibility and reliability on the international stage and within alliances. 84 Italian participation in both OEF and ISAF was an integral part of Italy’s use of the armed forces, as an instrument of foreign policy to improve prestige, to influence allies and, consequently, address other national security issues bilaterally with the US and within NATO. The subsequent paragraphs will now show how US officials applied ‘pressure’ through diplomats in Rome, and will elucidate how the Italian authorities reacted in order to improve Italian relative power bilaterally and within alliances. 85

In a report sent by the US Embassy in Rome on 15 September 2005 US Ambassador, Ronald Spogli, reported the likely handover from the centre-right government of Silvio Berlusconi to a centre-left government led by Romano Prodi. Spogli assessed the impact of such a centre-left coalition government after the 2006 elections on Italian military commitments in Iraq and Afghanistan. He argued that an Italian revision of the commitment to support the US Global War on Terror (GWOT) would not have any implications for Afghanistan, but would eventually affect the contingent in Iraq and, even in this case, the withdrawal would be coordinated with Washington. On this issue Spogli ensured the US leadership that he had obtained all necessary guarantees from officials of the centre-left Democratic Party and, therefore, such a change of leadership in Rome would not adversely affect the military presence in Afghanistan as endorsed by Berlusconi. 86

Spogli sent a second interesting report on 11 April 2008 to the US Secretary of State, Condoleezza Rice, again it reported on Italian national elections. Spogli highlighted the good relationship he had developed with the Prodi Government but that political leadership in Italy was about to change. The leadership race was likely to be between the centre-right candidate, Berlusconi, and the centre-left candidate, Walter Veltroni. According to Spogli both candidates offered continued support for US policy in the world, however he recommended direct engagement following the election whilst the new government was being formed, in order to influence the direction of Italian foreign policy. Spogli went further, stressing the importance of significant US official visits between April and May 2008 as the new government was forming to discuss important issues, including Iran, Russia, Syria, and above all, military commitment in Afghanistan. 87 Whatever the outcome of the elections Spogli stated his predilection

“... We will urge Italy to take greater ownership for its Command Region [RC-West]... We will press Italy to authorize its troops to take a more active approach against insurgent activity in Herat and Farah provinces and to ease or drop geographic caveats that prohibit Italian troops from crossing into RC-South... We will also encourage the Italians to do more Afghan police training, a field in which they excel, by expanding the Guardia di Finanza training to the Afghan Border Police, stepping up Carabinieri police training ..." 88

The evidence of the Italian Government acceptance of the US requests is self-evident in the subsequent US Ambassador reports. On 12 August 2008 Spogli communicated to US Vice President Dick Cheney, “Berlusconi relaxed geographic caveats for Italian ISAF forces, but we are monitoring progress on bringing promised police and border guard trainers into the field...”. 89 On 19 November 2008 Spogli
confirmed the success of US pressure on the Italian authorities to remove the national caveat on the regional employment of Italian troops in Afghanistan. Confident about the possibility of achieving more he suggested to Rice that during her visit to Rome, scheduled for 3 December, she should put pressure on Italian authorities to increase the number of troops deployed in Afghanistan, at that point 2,200. On 5 February 2009, in his final report before ending his mandate in Rome, Spogli encouraged the US President and Vice President to engage Italy on key issues (among these the Middle East and Afghanistan) and ensure her inclusion in the group of states that the US work with more closely to “…harness Italy's great potential in support of U.S. strategic goals”.

The US request to increase the Italian contingent in Afghanistan got the approval of the Minister of Foreign Affairs, Franco Frattini, and the Minister of Defence, Ignazio La Russa, but found strong opposition from the Minister of Finance, Giulio Tremonti. The latter adamantly opposed any increase of military expenditure for Afghanistan. It was Alessandro Ruben, La Russa's military advisor that told the new US Ambassador, David Thorne, (April 4, 2009) that through engagement at the highest levels of the US government Tremonti's resistance had been overcome. Frattini and La Russa's idea was of make the Italian surge in support of Afghan elections permanent instead of withdrawing them after August 2009. Regarding the issue of increased personnel to train the ANA, ANP and ANBP there would likely be more difficulties, but Ruben argued that were not insurmountable. On 9 June 2009 Thorne reported, through Hillary Clinton (the new US Secretary of State), to US president Barak Obama that Italy was keen to remove its national caveats, had increased her troop level to 2,600 and had doubled the number of deployed Carabinieri to one hundred, becoming the 6th largest ISAF contributor. If the force deployed for the Afghan election of August 2009 were to become permanent Italy would step “in the top tier of Afghan contributors”. Thorn added that "Berlusconi knows this is a priority area for the US and will likely respond positively if you [Obama] press him to do more in the region". The 9 June report is just one week before a meeting between Obama and Berlusconi. The report was entitled “Scene setter for Italian PM Berlusconi’s June 15 visit to Washington”. As a result of the meeting and, after having won over Tremonti, the Italian Government accepted an increase in its contingent. Thorne, in his report to Hillary Clinton on 22 January 2010, said that Italy had decided to increase her contingent in Afghanistan by one thousand troops during the first half of 2010. The information was confirmed on January 8, during a round table discussion by Stefano Ronca, the Director General for Multilateral Political Cooperation of the Ministry of Foreign Affairs. The 22 January report included additional important information. The head of the Foreign Policy Department of the opposition (centre-left coalition), Piero Fassino, stated: “the Democratic Party supports the increase of the contingent in Afghanistan”.

According to a survey published by James Sperling in 2012, Italy is an important European partner for the US although not on the same level as France, Germany and the UK. US Presidents (George W. Bush and Barak Obama) have emphasised the Italian contribution in Afghanistan, indeed over 60% of references were related to defence and security policies and the Italian contribution to NATO. Moreover, Italy rates a high profile in the US State and Defence Departments but is less well regarded in other departments, such as Agriculture and the Treasury Departments. This US emphasis on the Italian military role in Afghanistan and within NATO was, for the Italian authorities, proof that their foreign policy was creating the desired effects. Consequently, it empowered the policy on the use of armed forces across the political spectrum. So these selected documents evidence US ‘external’ pressure regarding the Afghan deployment and how Italy responded. The next step is to look at how this might relate to realist international theory.

Italy’s military involvement in Afghanistan, in support of both OEF and ISAF, were clearly the result of external drivers. Allies, especially the US, used their influence to drive up the size of the Italian contribution despite the lack of any economic advantages to Italy. In particular, pressure from the highest level of the US Government succeeded in overcoming domestic opposition and public opinion. The employment of Italian armed forces as an instrument of foreign policy was not dependant on any
particular party or coalition: both centre-right and centre-left governments were keen to continue showing their support. There was even bipartisan support for increases in troop numbers in Afghanistan, and for further guarantees of the Italian commitment to NATO-led operations alongside the US. Italy increasingly took on a leadership role; increasing the number of troops to become a top tier ISAF contributor by 2010, deciding to participate in OEF, the assumption of the lead in RC-West in 2006, the deployment of combat and privileged assets, (bombers, attack helicopters and UAVs), and the commitment to expand training of the ANA, ANP and ANBP. The aim was to increase relative power within NATO and bilaterally with the US. This increased influence was to directly offset any risk of Italy becoming marginalised in multinational organisations.

In his thoughts on alliances, Walt suggests that the influence of the dominant power is directly proportionate to the gap in resources between that power and other allies. Moreover, other great military powers, such as France and the UK, are clearly influential NATO members while Germany stands within the alliance, thanks to her economy. Therefore, in complying with US requests Italy tried to add more value to the alliance in order to improve her level of influence within it and bilaterally with the US.

**Italian participation in the multinational military intervention in Libya**

This case study will explain the factors that determined Italian military intervention in Libya in 2011. This support was provided in concert with allies in the aftermath of UNSCR 1973 approval, both under Operation Odyssey Dawn March 19-30 and subsequently in the NATO-led Operation Unified Protector, which started on 31 March 2011. Starting by setting the context (subsection I) from which the decision for intervention was taken, including the position of key allies, most notably the US, the UK and France and the strategic implications of operations in North Africa, in particular Libya. Second, it will identify those internal influences (subsection II), whose determinants, powerful economic actors and government coalition parties would have preferred to retain Colonel Muammar al-Gaddafi in power. Third, it will identify external factors (subsection III), which necessitated an increase in national relative power in comparison with allies and within the multinational organisations, most notably NATO, the UN and the EU. Finally, it will provide evidence to show that again external issues were the main determinants of Italian policy. The conclusions will make clear that military intervention in Libya was determined by external factors. The questions that this case study will strive to answer are: What motivated the Italian change of tack from backing a diplomatic solution to military action? Why did decision makers, notwithstanding internal pressure for supporting Gaddafi, decide to join allies in targeting Colonel Gaddafi's regime?

**Context**

On 17 March 2011 the UNSC authorised the “use of all necessary measures … to protect civilians and civilian populated areas under threat of attack in the Libyan Jamahiriya, including Benghazi”. This clause, included in paragraph 4 of UNSCR 1973, was interpreted to mean the UNSC approved the use of force against the Libyan armed forces being utilised by Colonel Gaddafi to suppress the popular uprising which had started in February 2011, and so threatening the survival of his regime. The UNSC approved resolution 1973 under Chapter 7 of the UN Charter to mitigate a threat to international peace and avoid a humanitarian crisis. The unusual factor in the Libyan crisis was the role of the US. France and then the UK asked for the implementation of a No Fly Zone (NFZ) over Libya, to stop the alleged uses of Libyan air power against protesters in Tripoli and other cities. US President Barak Obama and influential members of his administration, such as the Secretary of Defence Robert Gates, were reluctant to impose a NFZ which might be seen as another war in a Muslim country. Therefore, the US initially did not support the Franco-UK initiative. President Obama changed policy in support of military intervention just 48 hours before the approval of UNSCR 1973. This US change was due to regional support for a NFZ from the Arab League (March 12), the Gulf Cooperation Council (March 7) and the Organisation of the Islamic Conference (March 8). Regional support, multilateralism, and decisions
agreed with the majority of allies were the key pillars announced by President Obama in his ‘new’
defence strategy; the positive combination of these requirements explains the US shift towards use of
force.111

Gaddafi’s narrative to threaten to smash his own people into submission removed any qualms about a
NFZ, so the US asked and obtained UNSC authorisation to “use all necessary measures”.112 This US
‘policy pendulum’ has been highlighted because it impacted the Italian approach to the crisis and its
change of stance towards military action alongside allies. Coincidentally, influential EU countries, most
notably Germany, strongly disagreed with any policy for the use of force and the implementation of a
NFZ.113 The fact that Germany was a member of the UNSC in 2011 enhanced her position. Other UNSC
members also disagreed with the use of force; Russia and China, as P5 countries, but also India and
Brazil.114 The opposition of these countries within the UNSC, combined with US scepticism, reinforced
the Italian perception that the resolution would not be approved and, consequently, the likelihood of
any military intervention in Libya was assessed as low. Also complicating the issue was Italy’s long
standing political and historical links to North Africa.

North Africa, including the Sahel Region, West Africa, and the Horn of Africa, are areas of great interest
to Italy.115 In particular, North Africa has some strategic importance due to the economic opportunities
to import and export natural resources.116 However, not everything is positive for Italy in North Africa as
it is also a potential source of risks. Due to chronic socio-political instability, religious, ethnic tensions
and economic fragility North Africa affects regions well beyond its geographic border, in particular
within the closest European countries. The popular uprising known as the ‘Arab Spring’, which set
popular movements against the North African autocratic regimes, raised tension and concerns for the
instability of the region. Such instability retains the threat of radicalisation, the provision of safe havens
for terrorist groups, the smuggling of humans, weapons and drugs and, finally, through providing an
apparently endless flow of illegal immigrants. Human trafficking and illegal immigration are significant
issues for Italy.117

Libya is a ‘bottleneck’ for the illegal immigration of people from North Africa to Europe via Italy.118 Italy
and Libya are linked by a controversial history and, although recently Italian economic relations are
strong, illegal immigration has always been a considerable problem. Immigration flow from Libya
became a problem for Italy in the 1990s but, historically, it has been the exploitation of Africa’s natural
resources that has driven Italian policy in the region. By 2000 Italy had a unique bond and strong
economic links to Libya. This privileged position was the result of Gaddafi’s diplomatic isolation and the
US economic embargo, which had been established after the potential Libyan collusion with
international terrorism, of the 1980s and 1990s. After the 9/11 terrorist attack, Gaddafi started to
cooperate with the US’s Bush administration and it’s Global War on Terror. This rehabilitated Gaddafi
and his regime in the eyes of both the US and other western countries.119 With the removal of the
economic embargo and with the growing presence of eastern commercial competitors, most notably
China, India and Korea but also Turkey, the Italian ‘special relation’ with Libya was at stake. To mitigate
this risk national lobbies and powerful private economic actors, most notably the Italian multinational
energy giant ENI, started to apply pressure on the Italian government for a more active foreign policy
towards Libya.120 Moreover, tackling illegal immigration was a manifesto promise for the coalition’s
second biggest party, the Northern League (Lega Nord).121 To address these two core issues: illegal
immigration and the private businesses’ expectations for favourable access to the Libyan market and
natural resources, the Italian government signed the Treaty of Friendship with Libya.122 This paragraph
has set the context now it is important to address the key internal and external factors which drove
Italian decisions.
Internal factors

The Treaty of Friendship with Libya as a foreign policy course of action can be explained using all of the Moravcsik’s liberal theory variants: the commercial and republican liberalism, if we consider pressures applied by powerful private economic actors and political parties and the ideational liberalism, for public sensitivity towards border security adversely affected by the illegal immigration. But, more importantly, Moravcsik’s arguments give reason for the initial cautious Italian approach to the crisis, motivated by internal factors which required Colonel Gaddafi to remain in power and so guarantee the Treaty of Friendship and all the advantages it conferred. The scope and advantages ordered through the Treaty of Friendship were strong and did influence decision makers. The scope of both the security issue (regulation of migration) and economic implications (opportunities for the energy, air, naval and banking private sectors) addressed within the Treaty of Friendship will be summarised in order to highlight the range of national interests at stake and, in particular, how much stronger the domestic pressures were for supporting the Gaddafi regime rather than attacking it.

With the Treaty ratified by the Italian parliament on 3 February 2009, Italy agreed to pay $5 billion as compensation for her colonial past in Libya. This amount of money was to have been provided over 20 years but not in cash, rather through industrial projects realised by the Italian private sector operating in Libya. The main project, amount to $3 billion, was the construction of a motorway along the coast from the Libyan-Tunisian border to the Libyan-Egyptian border. In exchange, ENI signed contracts for the extraction of crude oil until 2042 and for natural gas until 2047. Libya exported 85% of her oil to Europe of which 14% was controlled by ENI. Overall Italy imported 23% of oil and 12% of its gas supplies from Libya. The Treaty allowed for gas provision to increase from 8 to 11 millions of cubic meters through the submarine gas line ‘Greestream’ built by ENI. The Treaty of Friendship gave a boost to the bilateral relations between the two countries and provided wide ranging economic opportunities for Italian industry: 10 AW 109 Power and AW 119 Koala helicopters were sold by Augusta-Westland; 4 maritime patrol aircraft ATR 42 MP and ATR 42 500 were sold by Alenia Aeronautica; the installation of a border control system by Selex Galileo, a Finmeccanica society; the creation of a joint venture between Finmeccanica and the Libya Africa Investment Portfolio, for investment across all African markets. Moreover, a robust influx of liquidity arrived in the banking system, mainly in UNICREDIT, one of the biggest Italian global banking and financial services company bank. The Treaty led to an overall increase in bilateral business volume of $13 billion but this was only part of the advantages.

In terms of illegal immigration to Europe from Libya, Italy provided six patrol boats, and Libya committed herself to seriously fighting the flow of African immigrants from her coast. After only 7 months the Italian Minister of the Interior, Roberto Maroni, officially declared that the flow of immigrants had reduced by 90%, reducing from 15,000 in the second semester of 2008, to just 1,400 in the second semester of 2009. This announcement was made on 9 September 2009 in Brdo, Slovenia, during the 19th Ministerial Conference on illegal immigration, organised crime, corruption and terrorism. These figures were attributable to the Treaty and, therefore, as a Government and in particular Lega Nord success. This reduction of landings reduced the international embarrassment caused by Italy's policy of immediate ‘forced’ repatriation. So the treaty was positive from several points of view. Finally, the treaty had a final clause which stated that Italy would not permit the use of bases on Italian soil for strikes against the Gaddafi regime. The respect of this clause was subordinate to the regime respect for international law. Therefore, the Italian authorities were unconcerned about its implications and how it would be perceived by Italy’s allies, most notably the US.

The Treaty of Friendship provided economic advantages to Italian businesses; however, the $5 billion of compensation would be paid mainly by Italian tax payers. The economic interests for private companies were huge and, therefore, great were the pressures applied on the decision makers to facilitate business transactions to the detriment of other foreign competitors. Another issue was the
inclusion in the treaty of immigrant controls. This was really sponsored by just one coalition party but, in this case, undoubtedly the benefit was not a lucrative one and not limited to a part of the population living in a specific geographic area. It was in these circumstances, of renewed relations between the two countries, within which the Libya crisis was dealt with by the Italian statesmen. The Italian government’s initial approach to the crisis will be now introduced. It will be explained as an attempt to preserve the status quo ante the blow-up of the crisis in Libya.

As the crisis began states and multinational organisations agreed to condemn the Gaddafi regime and his use of violence against civilians. On the other hand, no single course of action was agreed by the main actors on the international stage and within the UNSC. The United States in particular did not want to sanction the use of force, and its cautious behaviour and scepticism towards military action was coherent with the official non-intervention policy adopted in the ‘Arab Spring’ uprising. As remarked in the Afghan case study (section II), the US position on international issues and, in particular, those related to security have been taken into consideration by Italian decision makers since the end of the cold war, regardless of the coalition party governing at that time. The mention of international divergences, in particular those within the UNSC, serve the purpose of highlighting the lack of external pressures. Therefore, Italy’s initial approach to the crisis was mainly influenced by internal ‘forces’ aimed at a diplomatic solution to rehabilitated Gaddafi on the international stage, because he was key to ensure that contracts signed by private companies under the auspices of the Treaty of Friendship would be honoured. In this respect it is possible to interpret the speech by the Italian Foreign Affairs Minister, Franco Frattini, to the Foreign Affairs Joint Committees (3rd Foreign Affairs and Emigration committee of the Senate of the Republic and III Foreign and Community Affairs committee of the House of Deputies) delivered on 16 March 2011. Frattini called for an immediate cease fire between Gaddafi’s forces and the rebels, he stated that the Italian government disagreed with those who called for a NFZ, because Italy considered this would prove ineffective and that a diplomatic solution was the only option which might provide a long term solution. The Minister of Interior, Roberto Maroni, in the Interior Joint Committees (1st and 3rd committees of the Senate and I and III committees of the House of Deputies) echoed Frattini, stressing the impact of the potential refugees crisis as people fleeing the fighting arrived on the Italian coast as a consequence of violence, which would only get worse if other nations intervened militarily in Libya. The two ministers’ disagreement with the use of force also took into consideration public discontent for a possible Italian intervention.

In March 2011 only 40% of the Italian population supported military intervention and, in spite of this percentage rising in April to 52%, only 14% agreed on shelling military targets. Finally, any Italian military intervention through a “coalition of the willing” would have triggered parliamentarian debates around the incompatibility of such action with Article 11 of the Italian constitution. Therefore, a multiplicity of factors opposed any Italian participation in military action but the Italian approach changed as the situation escalated. The US change of tack towards the “use of all necessary measures” (March 15, 2011); the UK and France determination for the implementation of NFZ over Libya; the regional support of Arab states for the latter; the complete international isolation of Gaddafi regime and the approval of UNSCR 1973 (March 17, 2011) led to a reassessment by the Italian political authorities. The key driver was moving from internal to external factors.

External factors

Domestic pressure for backing Colonel Gaddaﬁ as the underwriter of the Treaty of Friendship were overshadowed by the necessity to stay alongside powerful allies, to increase Italy’s relative power against them and, therefore, improve the capacity to influence their choices.

Italy’s geographic proximity to Libya meant she could provide a natural platform for the alliance. The use of Italian airfields, air trafﬁc control and infrastructure meant that this was the place from which to mount combat air patrols to deliver a NFZ, and so a request to use Italian bases arrived in Rome almost
immediately. The request was no surprise as the Minister of Defence, Ignazio La Russa, before the approval of UNSCR 1973, had already listed the bases that might be made available to allies. These were Amendola, Gioia del Colle, Sigonella, Aviano, Trapani, Decimomannu and Pantelleria. Initially, during Operation Odyssey Dawn and the short term imposition of the NFZ in March 2011, Italian support was in terms of provision of air bases support to allies with naval and air assets (mainly command and control, intelligence, patrolling and surveillance activities). When NATO member states agreed to take over the military operation on 25 March, and with the change to Operation Unified Protector, additional pressure was applied by the allies for Italy to play a more active role along with those member states that were not participating in the raids. With a phone call on April 25, the Italian Prime Minister Silvio Berlusconi communicated to President Barack Obama the Italian decision to remove any national caveat which might limit Italian participation in Operation Unified Protector, this included “authorising air-to-ground strikes on Libyan regime targets”. President Obama expressed his “great appreciation for Italy’s many critical contributions to NATO-led coalition efforts to enforce UNSCR 1970 [imposition of an arms embargo] and UNSCR 1973, and acknowledged Prime Minister Berlusconi’s leadership on Libya”. Obama’s appreciation for this Italian commitment was exactly one of the effects at which such intervention was aimed. Yet the factors that changed Italian policy were somewhat more complex.

The first necessity for the Italian Prime Minister was to demonstrate loyalty with the US and NATO alliance. The North Atlantic alliance and its collective defence system under Article 5 of the Washington Treaty is the main security pillar against external threats. The US, as the only super power within the alliance, has often called for more commitment from European countries for NATO-led operations. Successive American administrations have spoken of the so called Free-Riders, who benefit from the NATO umbrella without providing a fair contribution. So the US has always applied diplomatic pressure for a more active approach from her allies. The supporting role assumed by the US in Operation Unified Protector was the result of the Obama’s National Security Strategy, and internal American factors which led to a clash between Obama and the US congress. Obama ordered the attack against Gaddafi military targets in Libya (Operation Odyssey Dawn, launched on 19 March 2011) without asking for authorisation from Congress and, as a consequence, the latter limited funding allocated for the operation. This meant that the US was unable to fully participate and so, after 9 days of operations, Obama decided that the US should provide a supporting role. This did not prevent the US President from using his influence with allies to request a more active role, stressing that Gaddafi was a direct threat to European security. Italy was not indifferent to the US request so she made military bases available and, on April 19, announced a limited deployment in Libya, alongside US, France and UK. This was to be a small number of military advisors to improve the rebels’ organisation and then finally removing all remaining caveats to the allocation of naval and air assets. It is no surprise that the first ally that Berlusconi called was Obama, to tell him the full extent of Italy’s commitment. This call was to stress that Italy was not only a ‘consumer’ of security provided by the US and NATO, but that Italy could and would actively contribute to it. Frattini justified the stronger military commitment in Libya as part of an international responsibility to intervene to avoid a humanitarian crisis, but soon after he added that “NATO and our allies strongly and repeatedly called for more flexible military support from Italy”. Hence, the necessity to avoid a mass killing was a necessary cause but insufficient to motivate the government commitment, even with the legitimacy provided by a UNSCR. It is worth noting that only eight NATO members, out of 28, participated in Operation Unified Protector and that after France and the UK, Italy carried out the highest number of combat sorties. Therefore, the military intervention was again mainly aimed at improving Italian influence towards the US and within NATO, and “avoid Italy’s international isolation... confirming Italy’s commitment to its Euro-Atlantic Bond”.

Improved influence, prestige and avoidance of the risk of isolation could explain all military intervention decisions in every context, but alongside this there was a second, more specific, concurrent factor. This was the need to contain the burgeoning political-military leadership of the new Franco-British axis.
The move of US focus to the Asia-Pacific region, combined with Germany’s lack of will to assume EU leadership, unless in the economic field, had created a vacuum of leadership that France and the UK were keen to fill. The Franco-British Defence and Security Treaty signed on 2 November 2009 was evidence of a bilateral strategic settlement to capitalise and combine military expenditure for the achievement of their national strategic objectives. The French President, Nicolas Sarkozy, was the first in the early stage of the crisis to recognise the rebels’ leadership, established as the National Transitional Council (NTC), as the solely legitimate representative of the Libyan population. France and UK provided diplomatic leadership to get approval of UNSCR 1973, combined with their lead role militarily in targeting Gaddafi regime. These actions combined to ensure that these two countries held a prestigious position: regionally towards the Arab League, locally towards the NTC and strategically within NATO. There was, therefore, for Italy a second requirement; that of rebalancing the power and, consequently, the influence that these two countries were able to exert jointly in Europe and NATO, but also at a local level in Libya.

The third issue that tipped the scales towards Italy joining allies in a military intervention was security. The instability created by the Libyan crisis would have created all the risks associated with failing or failed states, or where there is an ongoing civil war. Radicalisation and terrorist groups find fertile ground where there is a low level of security: they thrive on domestic grievances to obtain popular support. The potential implications of continuing strife in Libya were huge for Italy, where the proximity and the weakness of the Libyan border to Italian coasts threatened a return to mass illegal immigration. In addition, all the effort to address Mediterranean and North African’s security issues within the EU, NATO and UN frameworks would have been thwarted if Italy had stood by passively, or if Italy benefited from the crisis though opportunistic conduct. Therefore, Italy needed to participate to show the world that she was capable and credible on the international stage and, consequently, the stances proposed.

The fourth factor of influence was the approval of UNSCR 1973. It provided a legal framework for the use of force. At one stroke it eliminated any constitutional restraint on operations, including those raised by advocates of Article 11 of the Italian constitution. It also helped to reduce, but not completely eliminate, the unpopularity of the air strikes amongst the Italian people and mitigated any clash between the government and the centre-left opposition parties. Acting in compliance with international law and with a UNSCR were necessary requirements for Italian decision makers.

The President of the Italian Republic, Giorgio Napolitano, guarantor of the constitution, declared that the concessions for the use of Italian bases and the ‘more active role’ announced by the Prime Minister were consistent with the Italian constitution. The jurist Natalino Ronzitti argued that Italian participation and air strikes were legitimised from a constitutional stand point, because of the UNSCR 1973, approved under Chapter 7 of the UN charter as humanitarian intervention in support of peace. Moreover, the clause prohibiting the use of Italian bases for launching attacks against Libya, included in the Treaty of Friendship, was superseded by the Gaddafi’s numerous violations of both human rights and international law and by his non-compliance with the requirements of UNSCR 1970 and 1973. So, again, it was the international external aspects that determined Italian participation but also her level of commitment. In this regard the extent of the Italian contribution follows.

Operation Unified Protector lasted eight months, from April to November 2011; it was a joint air and maritime operation. Italy, France, the US and the UK deployed only mentors and advisors on the ground. Such participation alone was no guarantee that Italian’s objectives would be achieved. Therefore, the Italian contribution offered was to be ‘in line with’ that provided by the other two key actors. The subsequent table encapsulates the Italian combat assets allocated to operations in Libya. It allows these to be compared with the French and UK contributions.
This case study started by making it clear that the Italian government, based on internal pressures, coming from influential private economic actors and political groups, decided to establish bilateral security and economic agreements with the Gaddafi regime. In this respect Italy’s initial cautious behaviour in the early stage of the Libyan crisis can be explained by applying Moravcsik’s variants of the liberal theory. On the other hand Moravcsik’s arguments cannot explain the subsequent decision to ‘tear up’ the Treaty of Friendship with Libya and join the allies in bombing military targets and supporting rebels, whose aim was to overthrow the regime of the Colonel. Rather Krasner’s thoughts on states necessity to collaborate with each other, to mitigate risks associated with insecurity and uncertainty and, above all, Walt’s views on alliances and their implication for weaker states can give reasons for what motivated the Italian decision makers when decided to approve Italian participation in the multinational military operation in Libya.

**Conclusion**

This paper started by giving evidence on the importance of Italian participation in multinational military operations. Based upon the relevance assumed by the Italian military commitments abroad, in the
twenty-first century, it strived to explain whether decisions taken by policymakers were a reaction to international pressures or a result of domestic drivers. To answer such a question, this study used the two IR mainstream theories: realism and liberalism. The two case studies considered operations in Afghanistan (from 2001 to 2013) and in Libya (2011) demonstrated that the theory that better explains the behaviour of the Italian decision makers is realism, not least because decisions were a response to outside influences. These influences were, in some cases, so strong as to eclipse domestic drivers which criticised military commitments in countries, where there were not economic benefits or where a diplomatic solution would have improved private actors business.

On the whole, Italy’s decision to participate in OEF and ISAF was aimed at demonstrating that Italy could provide a key contribution to allies. OEF was a “coalition of willing” while ISAF was a NATO-led operation. For politicians and public opinion there rarely was a difference between the two military commitments. Both created the same internal opposition and concerns. Nevertheless, both were approved because of allies’ stimulus, and the necessity to give a boost to Italian influence within NATO, UN and bilaterally with powerful allies. The Italian military presence in Afghanistan has lasted longer than a decade. This has been possible thanks to bipartisan support that this author defined as the thread of Ariadne of the Italian foreign policy. The evidence provided made clear the strength of US influence, and how determined the Italian authorities were to increase national prestige bilaterally and internationally. In this regard it is possible to explain the decisions to contribute to the GWOT, launched by US President Bush in 2001, joining OEF (a combat operation) and those of becoming the leading RC-West nation and then of stepping up among the top tier of ISAF contributors.

Similarly, Italian support during Operation Odyssey Dawn in Libya, the concession to allow the use of Italian air force installations and Italian airspace, prompting further participation in the NATO-led Operation Unified Protector, were all determined by external factors and, therefore, explainable through the application of a realist approach. There was a strong desire to demonstrate to the US that Italy can add value in terms of power within NATO, and that the US can rely on Italy to contribute to overseas operations even if there are domestic interests that might suggest the opposite. Italian statesmen perceived the need to gain additional credit with the Obama administration, considering the US shift in emphasis to the Asia-Pacific region, thus proving Italian loyalty and raising the country up to a level of an important ally within Operation Unified Protector. This is the application of the concept of influence that Walt considers when he explains ‘why’ states ally and ‘what are’ the implications. Concurrently the Italian intervention was also aimed to contain other allies’ power across North Africa, locally in Libya, but also within NATO. This was especially true for France and the UK, who took the opportunity of the US participation being limited by financial restrictions to push forward new Franco-British leadership. Italy needed to rebalance this influence. The prestige gained in NATO, but also within the UN and the EU, was another important factor to consider as it would have empowered other issues raised by Italy at the international level such as security in the Mediterranean South region. Finally, an unstable Libya was not in Italy’s interest and, once again, only through participation could Italy have influence over decisions taken by allies.

The necessity to balance possible and unpredictable external threats (direct, indirect, conventional, unconventional, symmetric and asymmetric) is well understood by Italian politicians in the twenty-first century. To mitigate the lack of national resources Italian statesmen have demonstrated the reliance placed on alliances both bilaterally and within multinational organisations. The alliance with the US as the dominant power created two effects: an increased perception of security, but also a significant level of influence to which Italian policymaker are not indifferent. Italy’s governments are also concerned about risk of becoming marginalised across the Atlantic (US) and within Europe (with France, Germany and the UK). Afghanistan and Libya have provided opportunities through which Italy can show its resolve and commitment to its friends, through increasing its national relative power by using its armed forces to show this resolve.
Realist theory could explain other Italian decisions to use her armed forces. Italy has been a consistent supporter of allied military operations in the past decade. Nonetheless, as stated in the introduction, each case of military intervention is exclusive and, therefore, the realist theory cannot be assumed as the default ‘tool’ to explain any Italy’s decision to project her military power.

Finally, what is demonstrated in this paper has led to broader implications. Italy was motivated by the need to increase her relative power and national prestige. Would the lack of such a need, combined with the absence of a direct common threat, reduce the impact of external pressures to the benefit of domestic drivers? Could this explain the behavior of those members of NATO that are reluctant to provide a fair contribution to NATO-led operations?

Endnotes

3 Stefania Forte and Alessandro Marrrone, “L’Italia e le missioni internazionali” Documenti Istituto Affari internazionali (IAI) 12, no 5 (2012): 49
6 http://it.wikipedia.org/wiki/Militari_italiani_caduti_in_missioni_all’estero (accessed January 9, 2014). There are not official Italian Government sources reporting number of casualties in operations.
7 Mainly in the 1990s and associated with the personnel employed in the Balkans.
20 Ibid.
23 Ibid., 517.
24 Ibid., 518.
25 Ibid., 519-520.
26 Ibid., 525.
27 Moravcsik, “Taking Preferences Seriously”, 528
33. Tim Dunne and Brian Schmidt “Realism,” in *The Globalization of World Politics: An Introduction to International Relations*: 90.
35. Ibid.
43. Ibid., 111.
46. In the time frame 2001-2013 Italy changed 6 governments (alternating centre-left, centre-right coalitions and even government of technocrats) and six prime ministers (among this politicians of centre-left, centre-right and technocrats). See Italian Government, Presidenza Del Consiglio dei Ministri, http://www.governo.it/Governo/Governi/governi.html
49. Ibid.
50. Ibid., 1-2
52. Tricarico, “L’Italia e la NATO in Afghanistan un Approccio Integrato per la Stabilizzazione dell’Area,”: 2.
60. Parlamento Italiano, Leggi, Difesa e Forze Armate, Rifornimento Missioni,
63. Cottichia and Giacomiello, “All Together Now!”, 137.
68. Ibid.
69. Ibid.
71 Germano Dottori, “La Guerra (Quasi) Segreta degli Italiani”, 44.
72 Ibid.
75 Douglas M. Faherty, “Italian Foreign Policy: Trends For The Twenty-First Century”, 1
76 Ibid., 27.
77 Ibid., 29.
78 Ibid., 31.
79 Ibid.
80 Ibid., 33.
81 Davidson, “Italy-US Relationship since the End of the Cold War”; 295; Croci, “Italian foreign policy after the end of the cold war…”, 117.
84 Fabrizio Cottichia and Giampiero Giacomiello, “All Together Now!”, 145.
85 The subsequent information are selected from classified documents sent by the US ambassador in Rome to Washington, for a decade up to 2010. These documents were made available online by WikiLeaks, a journalist organisation. The ‘leaks’ of documents was interrupted in January 2011.
88 Ibid.
93 Ibid.
94 Ibid.
95 Ibid.
97 Ibid.
98 Ibid.
100 Ibid.
101 Ibid.
102 James Sperling, “Taken for granted or not taken seriously? American Perceptions of Italy as a Foreign Policy Actor” Istituto per gli Studi di Politica Internazionale (ISPI) no. 137 (2012): 1
103 Ibid., 3
104 Ibid.
105 Croci, “Italian foreign policy after the end of the cold war”, 127.
116 Ibid.
117 Ibid., 5.
118 Ibid.
120 Ibid.
121 Lombardi, “The Berlusconi Government and the Intervention in Libya”, 32
122 Varvelli, “Italy and Libya: Renewing a Special Relationship”, 125
124 Ibid.
125 Ibid.
126 Varvelli, “Italy and Libya: Renewing a Special Relationship”, 125
131 Ibid., “Italy-US Relationship since the End of the Cold War”, 289.
138 Ibid.
140 Ibid.
142 Ibid.
145 Ibid.
150 Giampiero Giacomello and Bertjan Verbeek, Italy’s Foreign Policy in the Twenty-First Century, 6.
151 Giovanni Bonvicini and Alessandro Colombo, La Politica Estera dell’Italia (Bologna, Il Mulino, 2011), 22; Italy, SMD DAS 004-1: 14.
The German forcing of the Meuse was the ‘inevitable outcome of the advantage the attacker has in choosing his point of attack’ (Alastair Horne, p468). Discuss this contemporary British assessment, especially with reference to how the French employed their mobile reserves.

Colonel J S A CARR-SMITH

“[On 15 May 1940] I was woken up with the news that [the French Prime Minister] M. Reynaud was on the telephone. He spoke in English, and evidently under stress. ‘We have been defeated.’ As I did not immediately respond he said again: ‘We are beaten; we have lost the battle.’ I said: ‘Surely it can’t have happened so soon?’ But he replied: ‘the front is broken near Sedan; they are pouring through in great numbers with tanks and armoured cars’.”

- Winston Churchill

INTRODUCTION

There was nothing inevitable about the German forcing of the River Meuse in May 1940 which resulted in the fall of France the following month. The Allies had several opportunities to disrupt the German forces and, arguably, even defeat them but they failed to exploit any of these, due to poor liaison, weak command and outdated doctrine. The Germans, on the other hand, particularly General Heinz Guderian, not only took advantage of Allied inactivity but also excelled in the areas of doctrine, command and Air-Land integration, not least due to a heavy concentration on the conceptual component of fighting power between the wars. Senior German commanders, including Guderian, spent their formative military years thinking, war-gaming and exercising without equipment and troops until they had a polished concept, and this emphasis on the conceptual component was to prove far more influential in the forcing of the Meuse than the Germans’ choosing precisely their point of attack. The fall of France in 1940 is an example that shows that an Army that is catastrophically dislocated or defeated is usually out-thought as much as it is out-fought.

Given Guderian’s association with German tactics, this paper will focus predominantly on him and his XIX Corps as a vehicle for analysis. Additionally, rather than focusing wholly on how the French employed their mobile reserves, this paper will analyse each of the opportunities the French had to both disrupt, and potentially defeat, the Germans during the period 10-15 May 1940. Finally, this paper examines the following three enduring themes that fall out of the Germans’ and Allies’ actions at Sedan in May 1940, that have relevance for conflict today and in the future: doctrine, command and Air-Land integration.

‘THE FRONT IS BROKEN NEAR SEDAN’

Background

In late 1939 the German Army High Command was considering how best to conduct an offensive in North West Europe. Their start point was to use the Schlieffen Plan of 1914, but General Manstein (Chief of Staff of the Army Group) and Guderian felt that this plan “had the advantage of simplicity, though hardly the charm of novelty”. Manstein, on the other hand, described by Guderian as “our finest operational brain”, did have some innovative thoughts and in November...
1939 summoned Guderian to discuss his plans. These involved a strong tank thrust through southern Belgium and Luxembourg towards Sedan, a break-through of the prolongation of the Maginot Line in that area, and a consequent splitting in two of the whole French front. Guderian, a tank proponent, was enthused by what he heard and told Manstein as such. Manstein then presented his plans to the German Army High Command. Whilst the plans were not initially universally well received, they were war-gamed at Koblenz on 7 February 1940 before becoming the formal operational design of the German Army, known as the Manstein Plan.

General Kleist was placed in charge of the operation, with Guderian's XIX Corps and Reinhardt's XLI Corps under command; the formation was known as Panzer Group Kleist. Guderian's XIX Corps formed the vanguard of the attack consisting of 1st, 2nd, and 10th Panzer Divisions, the Infantry Regiment Grossdeutschland and corps troops. Guderian knew all three of his divisional commanders well and had unbounded faith in their ability. These established relationships invariably greatly facilitated XIX Corps' preparations in late March and April 1940. As well as being pleased with his team, Guderian was confident of success ahead of the forthcoming operation. He felt that the French were: preoccupied with the concepts of positional warfare; too dependent on a complete picture of the enemy's order of battle before deciding how to react; and too reliant on planned control in which nothing should be left to chance. Put simply, Guderian was confident that the speed, manoeuvre and momentum of his XIX Panzer Corps, combined with German commanders who understood mission-orientated tactics (Auftragstaktik), would be too much of a match for the Allies. He felt that they, particularly the French, had swung from a strong emphasis on manoeuvre in 1914 to little-to-no emphasis on manoeuvre at the operational level in 1940, particularly in launching counter-attacks, and instead placed much greater emphasis on firepower and methodical battle.

The ‘impenetrable’ Ardennes

Once German preparations were complete the Manstein Plan began with XIX Corps crossing the border into Luxembourg on 10 May 1940, and advancing west through the Ardennes towards the River Meuse. The Belgians had responsibility amongst the Allies for this sector but, due to insufficient resources, they concentrated their forces on defending the centres of population and industry in the north, rather than the underpopulated wilderness of the Ardennes in the south. Consequently, the Belgians had only a light covering force in the south with orders that, should they be attacked, to withdraw northwards. They hoped that the French would arrive in time to secure the Ardennes. However, once attacked by XIX Corps the Belgians withdrew northwards, as instructed, but there was no hand-over to the French thereby leaving numerous obstacles uncovered. General Huntziger (Commander 2nd French Army) did eventually deploy a reconnaissance screen into the Ardennes but this was overwhelmed by 10 Panzer Division in the south, and what was left of Huntziger's cavalry withdrew west across the River Semois. Thus, due to poor liaison between Belgian and French forces, an opportunity was missed to delay XIX Corps' transit through the Ardennes whilst defences on the River Meuse were strengthened. Clearly, multinational alliances bring many strengths but they are not without their frictions; even when nations' existential interests are self-evident, cooperation and liaison can sometimes be more notional than actual.

Whilst XIX Corps did engage in combat with both Belgian and French forces, for instance at Bouillon, XIX Corps' principal challenge during the period 10-12 May 1940 was advancing along the steep and narrow roads through the Ardennes. This huge and extremely dense forest was described as both “impenetrable” and “Europe's best tank obstacle” by various French commanders. Nevertheless, the 41,000 vehicles belonging to Panzer Group Kleist did cross the Ardennes using only four transit routes. At one stage, the convoys traversing the narrow Ardennes roads extended 170 miles and would have been a prime target for enemy air forces, but this
opportunity could not be exploited fully by the Allies, not least because they did not have control of the air.

**Crossing of the River Meuse**

On 12 May 1940 XIX Corps crossed the River Semois, despite enemy artillery and air attacks hampering the construction of bridges throughout the day and, by nightfall, Guderian had seized the northern bank of the River Meuse. Kleist then instructed Guderian to cross the river at 1600 hours the next day. Guderian was nervous initially because, whilst 1st and 10th Panzer Divisions were firm on the north bank of the Meuse, 2nd Panzer Division had been delayed on the River Semois and would not be ready in time. Kleist was insistent, not least as he was keen to maintain momentum, and Guderian “felt obliged to admit that there were probably advantages in thrusting forwards immediately without waiting for all our troops to be ready”. It is interesting to note that Kleist, not Guderian, seemed the more eager to drive forward and seize the initiative in this instance, preferring to “attack at once, without wasting time so as to hit the French before they could regain their breath”. Only five days later Kleist nearly sacked Guderian for seeking to do exactly the same.

Despite having less time than he imagined to issue orders to cross the Meuse, Guderian and his staff were able to fall back on exactly the same orders used in the earlier war-games, all that was required was a change to the dates and times. The plan saw 2nd Panzer Division in the north forming the right flank of the assaulting force, 1st Panzer Division, reinforced by the Grossdeutschland Regiment and the divisional artillery of 2nd and 10th Panzer Divisions, in the centre forming the main assaulting force immediately north of Sedan, and 10th Panzer Division in the south forming the left flank. XIX Corps spent the night of 12 May preparing for their river assault crossing the next day.

Eventually, two days after Panzer Group Kleist began its advance west through the Ardennes, the French High Command reacted to events by ordering XXI Corps, under the command of General Flavigny, to reinforce Sedan to bolster X Corps, commanded by General Gransard. Much of X Corps consisted of Reservists; for example, the 55th Division was manned by fewer Regular officers and soldiers than any other French Division and yet it was the farthest left division in the Second Army charged with the defence of Sedan (the sector that Guderian’s XIX Corps crossed the afternoon of 13 May). Indeed Gransard later said: “discipline in these divisions [55th and 71st] was reduced by the slackness in the men and lack of leadership amongst most of the subalterns”. It was little surprise, therefore, that, despite being instructed on 12 May to reinforce Sedan, Flavigny’s XXI Corps did not react rapidly enough and did not arrive there until 14 May, after the German assault over the Meuse.

Nevertheless, the French were presented with yet another opportunity to disrupt German plans, this time on the morning of 13 May. Whilst Guderian’s XIX Corps was massing on the northern bank of the Meuse, Gransard reported that his X Corps could see the Germans “emerging from the forest… the deployment in general; everywhere observation points report an almost uninterrupted descent of infantrymen, of vehicles either armoured or motorised”. A French military critic, General Menu, wrote afterwards: “what an opportunity for the artillery to rain down those ‘hammer blows’”. Yet, despite the French doctrine at the time being based on positional warfare, so heavily reliant upon firepower, Gransard had instructed his men to limit their fire to 30 rounds per artillery piece. Such a rate of fire certainly did not constitute a ‘hammer blow’ and the window was missed, not least due to an overwhelming German Luftwaffe attack, the largest aerial assault in military history at that time.
Guderian understood not just that air attacks would be critical to the success of their advance but, importantly, that air and land forces must operate in a coordinated manner. Consequently, Luftwaffe commanders attended XIX Corps’ earlier war-games and Guderian himself attended Luftwaffe exercises. As a result, both Guderian and senior Luftwaffe commanders were keen to use air power in a new way: massed directed and guided artillery delivered in direct support to the advancing ground forces. This was a new departure for the Luftwaffe and a tactic that was to prove ultimately highly effective, providing further proof that the Germans had embraced fully the conceptual component of fighting power.

On the morning of 13 May Guderian visited 1st and 2nd Panzer Divisions to ensure that their preparations were running to plan, before going “to an advanced artillery O.P. of 10th Panzer Division in order personally to observe the effects of my artillery and of the Luftwaffe’s contribution”. The German air and artillery attacks on French forces on the southern bank of the Meuse started at 1200 hours on 13 May and lasted for over 4 hours, prior to the ground assault beginning at 1600 hours.

At 1600 hours XIX Corps’ infantry and assault engineers headed for the Meuse and began the initial crossing. Despite sustained air and artillery bombardments over the preceding four hours there was still some French resistance. Nevertheless, by nightfall XIX Corps had secured the far bank, the bunkers had been neutralised and the bridgehead could now be expanded. The success to date of the operation could be attributed, in no small part, to thorough preparation and training in the 6-8 weeks leading up to the attack, but also to Guderian’s personal drive, energy and leadership. Guderian spent the entire morning prior to the attack visiting his three divisional commanders, conducting face-to-face coordination, and explaining his aims for the upcoming operation. Furthermore, Guderian was routinely to be found at the point of main effort (Schwerpunkt) so that he could judge personally the state of the battle and thus re-allocate resources accordingly if required. As an example, Guderian crossed the Meuse in the first assault boat.

The French commanders’ situational awareness, on the other hand, and ability to impose their character on events as they unfolded were found wanting. To illustrate the point, although Guderian’s XIX Corps had secured the far bank of the Meuse, none of their tanks had crossed and XIX Corps was at its most vulnerable with only German infantry and assault engineers south of the river. Gransard instructed Lafontaine’s 55th Division to conduct a counter-attack at about 0430 hours on 14 May to destroy the German bridgehead, but none of the units were in position on time. The counter-attack did eventually take place at about 0700 hours but, by then, significant numbers of Guderian’s tanks had crossed the Meuse and the counter-attack was repelled. Lafontaine’s unwillingness to act immediately and decisively had allowed yet another opportunity to slip away.

Once the French counter-attack had failed, Guderian was able to expand his bridgehead and XIX Corps’ armour began to cross the Meuse in significant numbers. The French High Command realised how critical the bridges were to German success and, shortly after the failure of 55th Division’s counter-attack, the Allies launched an aerial attack. Between about 0900-1600 hours on 14 May British and French air bombers targeted the German bridges near Sedan but German ground air defence fires and the Luftwaffe managed to defend the bridges and, eventually, repel the attacks. The official British history notes: “no higher rate of loss in an operation of comparable size has ever been experienced by the Royal Air Force”.

The breakout

By midday on 14 May 1940 2nd Panzer Division, in the north, and 1st Panzer Division, in the centre, were ready to break-out of the bridgehead, but 10th Panzer Division, in the south, was facing
French resistance and experiencing difficulty getting its tanks across the Meuse. As such, the southern flank was exposed. Guderian, keen to maintain momentum, instructed 1st and 2nd Panzer Divisions to change direction, cross the Ardennes Canal and head west with the aim of breaking clear through the French defences. Meanwhile, the Grossdeutschland Regiment was ordered to guard XIX Corps’ southern flank in the vicinity of Stonne until 10th Panzer Division was across the Meuse and able to relieve them.

Flavigny’s XXI Corps, which had been instructed on 12 May to reinforce X Corps at Sedan, arrived eventually on 14 May. Flavigny’s orders were first to contain the Germans and then to counter-attack into the southern flank of XIX Corps at Stonne. XXI Corps consisted of 3rd Armoured Division, 3rd Motorised Infantry Division and the 5th Light Cavalry Division but, of the three divisions, only the motorized division was fully up to combat standards. For example, 3rd Armoured Division was only formed in March 1940; it had almost no maintenance or resupply capability; it had inadequate communications equipment; it had yet to receive its full complement of tanks; and it had never manoeuvred as a division due to a lack of training time.

Nevertheless, by about 1600 hours on 14 May, and with approximately four hours of daylight remaining, the bulk of XXI Corps was in position and “the tank crews were champing at the bit in their eagerness”. But Flavigny now made a fateful decision; he abandoned the counter-attack part of his orders in favour of the 1918 French principle of ‘containment first’. Consequently, 3rd Armoured Division and 3rd Motorised Infantry Division were instructed to disperse themselves across a broad front in a defensive posture. Rather than forming a strong force to counter-attack into the exposed southern flank of Guderian’s XIX Corps, and potentially deal them a catastrophic blow, the tanks of Flavigny’s XXI Corps were penny-packeted. As Colonel Goutard, commander of 3rd Armoured Division wrote afterwards: “there was a line, a few tanks but no 3rd Armoured Division. The steel lance was buried forever, and so was the counter-attack”. Under extreme pressure from General Georges, the Allied Commander-in-Chief North East Theatre, who was livid when he heard about XXI Corps’ defensive tactics, Flavigny did eventually pull together a counter-attack at 1730 hours on 15 May but it was a token effort and, no sooner had XXI Corps’ tanks entered into action the order was given to call off the attack. The Grossdeutschland Regiment did run into elements of Flavigny’s XXI Corps and heavy fighting ensued over the period 14-15 May; indeed, the village of Stonne changed hands several times, but 10th Panzer Division arrived just in time and Guderian’s southern flank was now secure.

The French had missed yet another opportunity to deal Guderian a devastating blow, arguably their best and certainly their last. General Hoth, Rommel’s Corps Commander, reckoned that as a result of the delay of Flavigny’s attack; “the French missed a favourable occasion; this counter-attack, conducted in a resolute manner, would have transformed defeat into victory”.

Guderian’s XIX Corps then advanced west and reached the English Channel on 21 May 1940 and cut-off the Allied troops from their rear area. Guderian had achieved in 11 days what the German Army in 1914 could not accomplish in four years. Fortunately for the Allies, Hitler issued a ‘Halt Order’ on 24 May 1940, resulting in the bulk of the British Army escaping from Dunkirk, many of whom would return to France later in the war.

“There are no Allies any more. There remains only one enemy: England!”

- Final Wehrmacht communiqué, 25 June 1940
ENDURING THEMES

When studying both the Germans’ and Allies’ actions at Sedan in May 1940, there are a number of enduring themes that have relevance for conflict today and in the future. This section of the paper looks briefly at the following three themes: doctrine, command and Air-Land integration.

Doctrine

France’s doctrine of positional warfare, with its inherent reliance on firepower, was no match for Guderian’s doctrine of speed, manoeuvre and momentum (known as Blitzkrieg). The French anticipated fighting a series of methodical battles but this had little chance of success against a highly mobile enemy, attacking deep into a defensive position complete with supporting fires. Despite a rich history of strategic and operational commanders and thinkers (e.g. Napoleon, Joffre and Foch) the French commanders in 1940 demonstrated scant regard for manoeuvre, particularly concerning counter-attacks. Guderian’s panzer divisions, on the other hand, not only outmanoeuvred the French, but their speed and momentum also surprised a number of German operational commanders including Kleist, Guderian’s superior. To illustrate the point, on 17 May 1940 Kleist ordered Guderian to halt as he felt consolidation to remove a potential threat on their flank was more important than exploiting the gains made by Guderian. Guderian objected strongly as he was keen to maintain momentum but Kleist stood his ground. Guderian therefore asked to be relieved of his command as a point of principle. Kleist reluctantly concurred but, after lobbying by other senior German commanders, Guderian was reinstated.

Turning to more recent times, it could be argued that the British Army’s way of thinking for much of the Cold War was more France 1940 (positional warfare) than Germany 1940 (manoeuvre warfare); i.e. spread-out to absorb the Warsaw Pact’s forces, provide a speed-bump and then threaten nuclear release. However, this changed when Field Marshal Bagnall grappled with NATO’s strategy of forward defence as Commander of the Northern Army Group in the mid-1980s, and introduced the concept of the operational level of war and manoeuvre warfare – i.e. the British Army finished the Cold War thinking more like Germany 1940 than France 1940.

In even more recent times, the opening attacks in the 2003 Iraq war, so-called shock and awe, is a modern-day equivalent of Guderian’s operational-level manoeuvre in May 1940 (Blitzkrieg), in that the US-led coalition began an intense and sustained air campaign on 21 March 2003 prior to coalition ground forces seizing Baghdad on 5 April 2003. In short, the coalition completely outmanoeuvred the Iraqi forces. The same could be said of the successful British and French interventions in Sierra Leone in 2000 and Mali in 2013 respectively. Although both campaigns were fought against rebel-groups rather than professional armies, both interventions were successful, at least in part, due to the British and French rapid responses combined with the use of devastating force when required.

Command

Guderian’s command style was undoubtedly a force multiplier. Whilst French commanders may well have been physically brave or even, on occasion, tactically proficient, they lacked Guderian’s intuitive flare (Fingerspitzengefühl) and deep appreciation of operational art. Guderian understood not just how to convert strategic objectives into tactical activity, but also how to employ, in this instance, an armoured force in an operational manner. He also favoured the principle of Auftragstaktik and sought to instil this into his commanders. He knew when to seize the moment and had the courage to say and act as such when challenged by superiors.
Additionally, Guderian understood the requirement not only to reinforce the main effort, as evidenced by pooling XIX Corps’ artillery assets with 1st Panzer Division as they crossed the Meuse, but also that to best influence the course of the battle a commander should be physically at the Schwerpunkt. Guderian’s command style was such that he rarely visited his Corps headquarters for more than an hour a day during the period 12-21 May 1940. He trusted his staff, and particularly his chief of staff, to coordinate activity in line with his intent. Instead, he favoured commanding from the front, as this enabled rapid decision-making and the opportunity to exert personal leadership to impose his will on his subordinates. It was Guderian’s understanding of operational art, and personal drive and determination to exploit it, which tipped the campaign fulcrum in Germany’s favour.

Many of Guderian’s qualities and insights are as relevant today as they were in 1940. Indeed, operational art, mission command, main effort and campaign fulcrum are all terms that modern day commanders would recognise. But in the modern era, principally due to bandwidth and connectivity, which can provide both clarity and ubiquity (e.g. UAV feeds), commanders, particularly senior commanders, often find that their situational awareness is vastly improved if they are located in their headquarters when critical decisions are required rather than physically at the point of main effort.

Additionally, senior modern-day commanders are as drawn to the political, social and cultural dynamics as they are to the purely military factors, and, again, these dynamics, which are routinely best experienced in cities and large towns, will often curb commanders’ natural instincts to predominantly be with their troops on the front line. Therefore, with the temptation to remain back in one’s headquarters with increased situational awareness being so irresistible, it is vital for commanders to know their subordinates really well in advance of the launch of operations. This is the only way to be in each other’s minds as the team collectively prosecutes operations using mission command. But this can be particularly challenging in multinational operations where, increasingly, commanders are presented with their subordinate commanders and formations shortly before an operation begins.

Clearly, it is a balance and there is no template. However, what is not in doubt is that Command is rightly viewed in this day and age as a capability in its own right rather than solely as a function of warfare.

**Air-Land integration**

Despite the opportunities that Allied air forces had to disrupt Guderian’s XIX Corps, Allied airpower in May 1940 was simply not as strong as German airpower and this, in part, goes some way to explaining why the Allies were unsuccessful in exploiting these opportunities. The Germans, on the other hand, not only enjoyed a superior air force at this stage of the war, but they also invested heavily in integrating the Luftwaffe with German ground forces. As has been mentioned, Luftwaffe commanders were closely involved in Guderian’s war-games prior to the Meuse crossing, and Guderian and other commanders joined the Luftwaffe in their planning exercises.

Trust between air and ground forces is a combat multiplier and only through habitual association and personal relationships can this trust be truly established. Guderian recognised this in North West Europe in 1940, as did Field Marshal Slim in Burma in 1942-1945, and General Franks in Iraq in 2003. All three commanders not only preached how important Air-Land integration was to the success of their campaigns, but they also led by example. To illustrate the point, Slim, Air Marshal Baldwin (who commanded Third Tactical Air Force) and Brigadier General Old (commander of the Anglo-American Troop Carrier Command), pooled intelligence resources in the Burma campaign, their planners worked together, and the commanders and principal staff lived in the same mess. In
effect, they formed a joint headquarters. There are no single-service solutions to modern-day conflict.

CONCLUSION

The outcome of the battle either side of the River Meuse in May 1940 was never a foregone conclusion. The Allies, particularly the French, had numerous opportunities to disrupt Guderian’s XIX Corps and, arguably, even defeat it, but they failed to exploit them due to poor liaison, weak command and outdated doctrine. The missed opportunities included: the weak Belgian screen and shoddy hand-off to the French cavalry in the Ardennes; Guderian’s XIX Corps massing on the northern bank of the Meuse prior to the crossing presenting an attractive target to Allied air and artillery; Guderian’s infantry and assault engineers forming a bridgehead immediately following the crossing was rife for destruction by a strong French counter-attack; XIX Corps bridges over the Meuse were critical to the success of the operation but none were destroyed by Allied air or artillery; and finally XIX Corps’ southern flank was exposed during the breakout, but once again, the French failed to conduct the counter-attack which many commentators say could have turned the tide in France’s favour.

But to place the blame entirely on Allied failures runs the risk of underplaying Guderian’s role. Guderian unquestionably understood the importance of the conceptual component of fighting power, and that to defeat the Allies the Germans would have to out-think them as much as out-fight them. Ultimately, it was Guderian’s understanding of operational art (converting “strategic objectives into tactical activity in order to achieve a desired outcome”), and personal drive and determination to exploit it, that was to prove far more influential in the forcing of the Meuse than the Germans’ choosing precisely their point of attack.

Guderian’s bold and decisive moves undoubtedly presented the Allies with opportunities, but the tempo of XIX Corps was so rapid that the French were simply over-run. Guderian’s intuitive flare (Fingerspitzengefühl), understanding of when to take the initiative via mission-orientated tactics (Auftragstaktik), and presence at the point of main effort (Schwerpunkt) to best influence the battle during key moments, were qualities found lacking amongst French leaders either side of the Meuse, and contributed to the fall of France in June 1940.

Endnotes

1 Alistair Horne, To Lose a Battle: France 1940 (Harmondsworth: Penguin, 1979), 433.
3 Ibid., 90.
4 Ibid., 89.
5 Ibid., 91.
6 Ibid., 94-96.
8 Horne, To Lose a Battle, 229.
9 Ibid.
12 Krause and Phillips, Operational Art, 171.
14 Guderian, Panzer Leader, 101.
15 Horne, To Lose a Battle, 305.
16 Guderian, Panzer Leader, 101.
17 Horne, To Lose a Battle, 322.
18 Krause and Phillips, Operational Art, 95.
20 Horne, *To Lose a Battle*, 324.
21 Ibid., 325.
22 Ibid.
25 Ibid., 102.
29 Ibid.
34 Ibid., 283-284.
35 Horne, *To Lose a Battle*, 386.
36 Ibid.
37 Ibid., 387.
38 Ibid., 411.
42 For more on Hitler’s ‘Halt Order’ see Horne, *To Lose a Battle*, 598.
43 Horne, *To Lose a Battle*, 646.
50 William Slim, *Defeat into Victory* (London: Pan, 2009), 243-244.
51 UK. DCDC. *Campaigning*. 3-6.
'Mid-May 1940 was an unimaginable hell for British and French forces alike'. As the experience of the French and German forces in 1940 demonstrates, doctrine, not equipment, is the key to success in modern warfare. Discuss.

COLONEL S SCHEIBL GAF

Introduction

‘Within the limits of its size, an army’s worth as a military instrument equals the quality and quantity of its equipment multiplied by its fighting power.”

Martin van Creveld

On 22 June 1940, France surrendered just six weeks after Germany launched its offensive against the Allies. After the defeat, a search for the culprits began in France. Who and what were responsible for France’s catastrophe? The experience of murderous defensive battles in Verdun or at the Somme dated back 25 years. Then France had built hundreds of bunkers along the Maginot Line and protected its east border in concrete. What the war power of imperial Germany had failed to achieve after decades of preparation in 1914 now succeeded in a matter of weeks. How could the actions of arch-enemy decide the war so fast? Perhaps one of the first to comment was the Supreme Commander French Land Forces, General Maurice Gamelin who, on 16 May after the German breakthrough at Sedan, when asked by Churchill about the counter-attack only replied: “Inferiority of numbers, inferiority of equipment, inferiority of method.”

This study looks at the reasons for France’s sudden and dramatic collapse. Mindful of van Creveld’s formula, it will look at the quantity and quality of the equipment, methods, and will look at determinants of fighting power such as morale and initiative, the willingness to fight and the readiness, if necessary, to die. The study reveals that doctrine, as a set of beliefs, played an important role, but was not the decisive cause for defeat and victory.

Numbers and equipment

The impression of defeat, in Gamelin’s judgement, is understandable but an analysis of the figures showed that Gamelin was wrong about French numbers. Various writers point out that the general strength ratio between the Allies’ and Germany’s equipment was actually in favour of the Allies. Frieser, for example, lists 135 German divisions (three million soldiers), who faced a total of 151 Allied divisions (four million soldiers). On 10 May 1940, the Allies outnumbered the German in artillery 14,000 to 7,378, in battle tanks 4,204 to 2,439 and in combat aircraft 4,469 to 3,578. Quantity was not counteracted by quality, quite the reverse. Kirkland stated that the French planes were comparable in combat capability and performance to the German aircraft. The newest French and British fighters Dewoitine 520 or Spitfire were, in some ways, superior to the Me 109. The other backbone of the German air force, the Ju-87 ‘Stuka’ dive-bomber, was poorly armoured, slow and highly vulnerable. The qualitative comparison of battle tanks further increases the imbalance of the strength ratio. The French battle tanks were decisively better armoured and the armament of the tanks were considered to be the best. Two thirds of the German battle tanks, Panzer I and II, were light-armoured and under-gunned. The more valuable and capable Panzer III, especially the Mark IV, were new and available in only small numbers. Furthermore, only a small
percentage of the army consisted of very modern armoured divisions. Most of the ground forces took part in the ‘Blitzkrieg’ with old material and slow horse-carriages. Only the anti-aircraft artillery strength ratio was strongly in favour of Germany regarding quantity and quality. There had to be different factors that determined the success of the German forces.

By 1939 most of the French bombers and tanks, in contrast to German assets, were not equipped with radios, which had a severe impact on operations. Being en route, there was no change in plans possible. Even French headquarters were not equipped with wireless communication, resulting in delays regarding the issuing of orders and the reception of reports, when the telephone lines were interrupted. Equipping the German tanks with radios did not mean that they become more controlled in terms of obedience. The effect was an increasing situational awareness for the tank crews and headquarters; both could obtain a rapid insight into what was going on resulting in a higher flexibility.

Other important criteria for the worth of a material are the degree of reliance and, regarding the use of aircraft, the sortie rate (how often can an aircraft be launched a day). Taking into account that 2,300 of the 2,900 French fighter aircraft and all of the 400 assault bombers were capable of penetrating the topside of German tanks with their 20mm-gun, only forty-eight anti-tank sorties were flown over a four day period in June without any considerable effect. The reason, beside a lack of armour-piercing ammunition, was an extremely low sortie-rate. The rate of many units was less than one sortie per aircraft per day (e.g. RAF day bomber-units 0.25, French reconnaissance-units 0.4). The operational rate of the French fighter force was 0.9 at the height of the battle. German fighter units flew up to four sorties per aircraft per day. So the high numbers of modern French aircraft on the ground did not really matter. The low sortie-rates caused by incomplete infrastructure and ground equipment, and coupled with insufficient personnel to man and serve the aircraft, are a definite indication that the French air force was not well prepared for war. How badly the organizational structure affected the overall-performance of the French air force becomes apparent when one considers the kill-loss ratio. When airborne, the French fighter pilots and aircraft proved to be more effective than their adversaries.

Regarding the later considerations it is important to have a closer look into the detailed numbers. From January 1939, each month between 370 and 600 aircraft were delivered to the French air force by the industry. The French Chief of Air Force, General Vuillemin, argued in 1939 that he could only handle 40 to 60 due to the lack of air and ground crews, who had to retire due to military-political disputes. Falling back on forty- to forty-five-year-old reservists to fly state-of-the-art airplanes, he finally accepted 330 aircraft per month. Kirkland pointed out that “The availability of aircrews and the physical capacities of the ageing pilots became the limiting factors on how frequently the aircraft could fly.” Despite the loss of 892 aircraft, by the armistice the French air force was stronger and better equipped than at the start of the German offensive. Taking into consideration that, beside 384 British operational combat aircraft on the Western front as of 10 May 1940, 1150 British aircraft were available at the homeland, the use of this immense force should have had another result than defeat. Furthermore, German industry could not keep up with the Allies’ production rates. The French air force received between 10 May and 12 June another 1,131 aircraft. These figures show that the noted low sortie rate was caused by a late force build-up: the air force structure and logistics could not keep pace with the industrial production rate of aircraft. Germany benefited from a higher operational readiness state in comparison to the Allies on 10 May 1940.

The worth of equipment is, besides the quantity and quality, determined by its use, or more precisely, when, where and how it is put in action. Discussion of these issues has to recognize that many factors are closely linked to other issues like doctrine or strategy. A good example is the Ju-87 ‘Stuka’. The decision to purchase the dive-bomber was an outcome of the Versailles treaty.
restrictions. Not earlier than 1935, when Hitler ignored the restrictions and established the German air force, the mass production of aircraft started. Because of Hitler’s timeline to enter war, Germany had to concentrate on light and medium bombers instead of the production of heavy bombers and sophisticated fighters requiring larger quantities of labour and material. Furthermore, facing the wars that Germany intended to fight, the function of the German air force was to support the army. Despite afore mentioned disadvantages, the Stuka had two qualities which affected the course of events. Based on the diving technique it could place a bomb on a target with greater precision than a bomber delivering out of level flight, the prevalent technique at the time. One of the factors associated with success was the use of the Stuka at the battlefield attacking the French defence line before the German artillery could reach the battlefield. These units had lower priority in the deployment phase and the quick advance through the narrows of the Eifel and the Ardennes: “What the Wehrmacht lacked in guns on the ground,” Horne concluded, “was more than made up for by its ‘flying artillery’.”

But it was not the Stuka’s bombing capabilities alone that ensured success. Actually, the 1,215 bomber and dive-bomber sorties flown in a 3-miles-wide sector inside and south of the Meuse River Loop at Sedan caused the destruction of only two bunkers by a direct hit. The side-effects of the twelve-hour rolling-raid against French positions on the Meuse on 13 May by 850 fighter-bombers, especially the Stuka, were reasons for success. The command of the French 55th Infantry Division was paralyzed because all the telecommunication cables that were frequently laid out in the open had been cut. In the German and French after action reports, the raid was rated as decisive in its desired psychological effect. Even if the bombs did not destroy the thick concrete walls, the continuous bombing had a psychological effect. French soldiers were paralyzed in their bunkers by the uninterrupted detonations and the shrill sirens of the diving bombers. The demoralization was described by a General as “Five hours of this nightmare was enough to shatter their nerves, and they became incapable of reacting against the enemy infantry.” Referring to Liddell Hart’s ‘indirect approach’, Frieser concluded that the Blitzkrieg “… was directed not against the muscles, but rather against the nerves of the enemy.” Avoiding combat action whenever it was possible, the German tank formations advanced deep into the enemy’s rear areas spreading chaos; the French front collapsed by itself. “The principle of psychological confusion replaced the old principle of physical annihilation.”

Operational methods and doctrine

The plan and the methods to attack France were based on doctrinal concepts developed by the first head of the interwar German army, General Hans von Seeckt. WW 1 experiences of infiltration tactics and successful battles at the East-front resulted in stressing penetration, breakthrough and manoeuvre instead of firepower to destruct the enemy. The aim was to target the weak spot of the enemy’s resistance, rupturing his defence. Parallel to the development of German field regulations, a general staff officer, Heinz Guderian, analysed the reasons for defeat in 1919 and recognized the potential of modern technology that is a fully mechanized force. His ideas evolved over a timeframe of fifteen years and were published 1937 in a book entitled “Achtung – Panzer!” Having learned the lesson that a war fought in trenches results in a long war with inadmissible losses, he pointed out how vital the proper use of tanks and supporting armoured vehicles, operating alone and not in support of the infantry, would be in the future conduct of war. To fight a mobile and flexible war demanded the solution of questions regarding supply (large fuel tanks, the use of jerry cans) or protection. Knowing the disadvantages of a slow infantry and the weakness of his concept, operating without the protection of the artillery (self-propelled guns where at this time an exception), Guderian saw the necessity to work closely with a tactical air force, to delay the movement of the enemy’s tanks and reserves. Furthermore, the support of the air force would counterattack the potential threat by the enemy’s fighter-bomber. Air superiority would guarantee protection and freedom to manoeuvre. But to fight a mobile and flexible war demands an
appropriate command and leadership. Here, Guderian’s ideas were congenial to the German doctrine emphasizing decentralization and initiative. “Independence of action of the lower commanders … is of decisive importance all times.” The philosophy of command called “Auftragstaktik” (mission-oriented command) was, and still is, that subordinate leaders or even common soldiers understand not only their commander’s intent, but also the intent of the next higher commander, and act accordingly even violating other guidance or orders if necessary. The German instruction to the commanders to hold an area, instead of holding a line, gave them the opportunity to move according the circumstances in the assigned area: “… understanding of the circumstances … is the crucial factor in the battle.” This is of even more vital importance in highly manoeuvrable warfare. So Guderian’s ideas fell on fertile ground.

Similar to the German military, the French also started a process after the end of WW I to gain the appropriate doctrine for the next war. The result was a doctrine which strongly emphasized the defensive and, based on technical development, firepower. The enormous firepower of antitank weapons and artillery would discourage an attacker. Conversely, the French believed that a successful attack would require a much larger number of troops and material. As a result, the French developed a step-by-step approach called ‘methodical battle’, where the weapon deployment and the movement of all units were tightly controlled by a schedule. This approach assumed a centralized command and control system, stressing hierarchy and, consequentially, obedience. Reinforced by the denial of radios, as a means for command, the method was inflexible and prone to surprises and changes on the battlefield. As the conduct of events in May and June 1940 showed, German doctrine emphasizing mobility, decentralization and initiative was appropriate for the ‘Blitz-operations’. The French doctrine stressing methodical battles, centralization and obedience, could not cope with this kind of war. However, studying the Polish campaign the French command realized the value of large armoured formations and hastily formed three tank divisions. But in May 1940 they were not completely equipped, rarely trained and the doctrine was not adopted yet. The bulk of the French tank force was still collocated with the infantry. There was no concentration of force, and the tanks had to stick with those slow units. It was not only the French army which did not have the appropriate ways and means to fight this war. Until 1936 the French air force had no aircraft suitable for the attack of battlefield targets, because the air staff on several occasions declined to consider proposals. The attack of battlefield targets was contrary to policy. The majority of the French aircraft were bombers designed for medium level bombing, of strategic targets. Forced to treetop-flying by the effective German anti-aircraft artillery, crews dropped their bombs without precise identification of tactical targets and experienced high losses. So-called ‘Provisional instructions for rapid air support of ground operations’ were rushed out on 30 May, but it was too late. Opposing, the engagement of the German air force in the Spanish civil war and the experiences gained in the war against Poland had immediate impact on techniques and tactics. That the French forces had to fight battle-trained troops was surely another reason for their defeat.

**Strategy and the national environment**

Preparing for another conflict with Germany, France developed a defensive strategy on the basis of the experiences made in WW I. On the assumption that Germany would repeat its 1914 Schlieffen-Plan, attacking through northern Belgium, the plan was to defend the territory by a massive line of fortification along the northeaster border and, after deploying troops into Belgium along the river Dyle, defending against the Germans as long as sufficient forces of the Allies were mobilized for a counterattack. Furthermore, the forested and hilly region opposing the Maginot Line and the river Meuse were considered natural obstacles for the deployment of larger particularly mechanized formations. Geography and fortification would, on one hand, force Germany to divert its main attack through Belgium and would, on the other hand, allow the concentration of the majority of the divisions in the northwest. Indeed, after the terrible losses in
WW I, France lacked access to manpower of the critical age of between 20 and 30. Furthermore, the French divisions were not all of the same quality. Therefore, the stronger divisions were supposed to fight the German troops on Belgian territory while second-line divisions were deployed along the Maginot Line.

While France’s objective was to avoid defeat, the German strategy was to achieve a fast and decisive victory. Germany was aware that, due to its limited industrial capabilities, it could not win long-lasting trench warfare like 1914-1919. The German general staff realized that victory could only be achieved by bypassing the Allies main defences in Belgium. Therefore, the German strategy in the 1940 campaign was to launch an attack through the Netherlands and Belgium to coerce the Allies to move their forces north. But the surprising and decisive element of the plan was a simultaneous attack with large mechanized formations through the Ardennes, battering the Maginot Line to achieve a breakthrough at the river Meuse in order to defeat French and British forces in the northwest.

The course of events played into the Germans’ hands. After the invasion of the Netherlands and Belgium, Gamelin, still assuming that this was the main attack, directed the designated forces and the reserves towards Belgium weakening the Maginot Line. His decision to move the reserve was decisive for the outcome of the war. French reserves misspent, German troops ruptured the Allied front at Sedan by the evening of 15 May. The German Commander-in-Chief Army Group B, General von Bock stated after the breakthrough at Sedan; “The French really do appear to have taken leave of their senses; otherwise they could and must prevent it.” French defences soon collapsed and on 22 June 1940, just six weeks after the invasion, France capitulated.

In retrospect, two British decisions have to be considered as strategic flaws too. Fighter Command held back numerous squadrons for the Battle of Britain instead to deny Germany’s air superiority and air-to-ground campaign in France. Instead of bombing troops and airfields, Bomber Command started on 15 May his unsuccessfully ‘private war’ against oil targets in Germany aimed at hampering the war-machine, and forcing Germany to withdraw forces from the front to protect its territory. Both decisions neglected an accurate assessment of enemy force capabilities and misjudged enemy’s intent. Germany did not withdraw any forces, but was willing to take casualties and damage in order to reach its objectives.

Despite the two obvious flaws, assuming a repetition of the German WWI plan and depriving any reply by sacrificing the reserves, the Vichy-Regime attempted to apportion blame on the failure of earlier French governments to protect the military. And indeed the circumstances of the collapse would suggest alternate explanations. The terrible war losses, the bad economic situation, the immense public debt caused by the war, and a very influential left and peace-minded intellectual scene were the setup for governmental instability. The left and right fought against each other by means and ways which were close to the “Beginnings of civil war.” There is some evidence that the events and the atmosphere of the inter-war years had some impact on the morale of the soldiers. But the analysis of the six-week war showed although that the French Army suffered a much higher rate of casualties: “… the French resisted to the last man” as a German military diarist wrote.

However, the French defensive plan was neither the result of the military assessment of German power nor of anti-militarism. It was rather the result of a process which started directly after victory in 1919. Still under the impression of trench warfare, a discussion began amongst the French commanders about the potential of fortifications. One school argued that fortifications bind material (mainly money) and personnel resources at the expense of other plans. The other school argued that fortifications do not have a defensive character only but enable the release of forces for offensive operations. In the end, the defensive orientation of the forces was linked to the
government’s decision in 1929 to reduce the length of the conscription to one year. For the French army it was impossible to execute offensive operations with short-term conscripts. These could only fight in a methodically way and could not handle new technology or new methods of warfare. In her analysis of the cultural context within the decisions were made, Elisabeth Kier concluded a conceptual barrier stood in the way of an offensive orientation after 1929. “The French army had the money, ideas and freedom to adopt an offensive doctrine, but instead chose a defensive doctrine. Its organizational culture would not allow otherwise.”

That conceptual barrier became clearer when the French command ignored intelligence reports predicting the attack near Sedan or reports from air reconnaissance about endless columns in the Ardennes on 12 May. French commanders could not imagine the audacity of the German offensive and missed the chance for an aerial counterattack. The German plan was successful because the French were not able to anticipate a surprising and risky plan. Within this cultural context, two differences in attitudes come to light. The noted lack of radios might be linked to the old age of the French generals. Many held high command in WW I and were at least sixty years old. Certain adverseness against technical aspects was obvious. Gamelin admitted after the war: “What would we, at this level, have done with a transmitter?”

German officers led using radio equipped command vehicles. The pictures of Guderian and others taken at the front during face-to-face briefings with front-line commanders are countless and symptomatic. The French general officers were trapped in methodical leadership principles and in the words of Marc Bloch: “… incapable of thinking in terms of new war.”

Conclusion

Germany, outnumbered in men and material, defeated the Allies 1940. Investments regarding the quality of the material, like radios, were made to strengthen the worth of its force, but greater efforts were undertaken to qualify personnel by training and education resulting in an evolution of operational methods and doctrine. The social and political situations were enabling ideas and solutions in direct support of the bedevilled intents of the German regime. Referring to van Creveld’s formula, all – doctrine, training, command technique, organization and administration – was “geared to fighting.” The equation for the French army was not miscalculated. The quantity and quality of soldiers and material were sufficient, the doctrine was appropriate for the plan, but not for the war France had to fight. So Gamelin was wrong about the inferiority of numbers and quality, but he was right regarding the inferiority of method. German forces honoured principles or factors of warfare, which are associated with success, like unity of effort and concentration of force, to a much greater extent. The assault through the Ardennes in a battering attack against the weak point of the enemy was a truly joint campaign. Furthermore, German forces demonstrated initiative and will in unauthorized risky actions like the breakout from the Sedan bridgehead. France and the Allies disregarded an overall joint command within their national and over all deployed forces. There was no joint warfare and no realistic assessment of friendly and enemy force capabilities. Germany’s intent and objectives were misjudged. And at last it has to be said; France lacked a share of good fortune. Or in the words of a British intelligence officer: “The Germans have taken every risk – criminally foolish risks – and they have got away with it.”

But when the irresistible ‘Blitzkrieg-doctrine’ turned out to be a myth, it was not bad luck. That, being successful against weaker, badly equipped or unprepared opponents, failed in later campaigns. In Russia, the German troops were confronted with the immensity of the steppes and could not cover them at all: “… they lacked density.” The Allies, like Montgomery in North Africa, developed their own operational methods reducing the battle space for German manoeuvrable warfare, and eventually succeeded. “It was victory that gave Blitzkrieg the status of doctrine.”

So what are the enduring lessons of a campaign, when entire nations went to war? In post-heroic societies losses will be not accepted. The footprint (quantity), even of all-volunteer forces, is developing increasingly smaller. Declining defence budgets will reinforce this trend. Technological
advantage (quality) over potential and real enemies has protected our soldiers in combat in the last decades but is declining. It was argued that France, which had assumed that the next war will be fought, according the same rules like the last war, was not able to fight that kind of a new war. It became obvious that France and its Allies failed to cope with the manoeuvrable warfare. As contemporary operations like Iraq and Afghanistan have shown, the character of operations does not change from conflict to conflict, from war to war; it is rapidly involving within an operation. The knowledge of our enemies, an accurate assessment of their capabilities, human intelligence or information superiority becomes more and more vital. Operational thinking and operational methods are already heavily context-dependent. To understand the circumstances and to cope with the challenges and threats ahead, “our ability to remain adaptable for the future will be fundamental.”

This becomes doctrine by itself.

Endnotes

4 Kirkland, op. cit., 102.
5 Frieser, op. cit., 38-42.
6 Horne, op. cit., 102 and 239.
7 Kirkland, op. cit., 110-11 and 115. In comparison to the French technical support and supply capabilities it has to be considered that most of German’s air force technical units were not operating from the home bases only, but although from occupied or auxiliary airfields.
8 According Kirkland the kill-loss ratio of the different fighter squadrons were between 2:1 and 4:1. ibid., 109-10.
9 ibid., 108.
10 Horne, op. cit., 159.
11 Frieser, op. cit., 158.
12 The British Major General Lord Trenchard, in view of the bombing raids during WW I, had already asserted that the effect on morale was in ration of 20:1 to the material effect. In: Frieser, op. cit., 345.
13 The howling screech during the diving delivery of the bombs was caused by a technical trick, a wing-fixed organ pipe, the so-called ‘trumpet of Jericho’.
14 Cited in: Horne, op. cit., 249. Airstrikes, not been integrated in a combined attack from ground forces, will not have a long lasting effect on the psyche of soldiers; they will learn to live with. See Mark Laity, “The latest Test for NATO” The RUSI Journal Feb/Mar Vol.157, No. 1 (2012): 56.
15 Frieser, op. cit., 345.
16 Frieser, op. cit., 344.
19 The latest ideas of the High Command were described in the French Revue d’ Infanterie Dec 1938: “Not even the most modern tanks can ever lead the fighting by themselves and for themselves. Their mission must always participate along with the fire of the artillery and heavy arms in the protection and the support of attacks.” Cited in: William L Shirer, The Collapse of the Third Republic (London: The Literary Guild, 1970), 159.
20 The long struggle of the French aviators for independence ended in a single service but although in a severe quarrel between Air Force and Army.
22 Neither the Netherlands nor Belgium were willing to enter into detailed military cooperation with their Allies (incl. the pre-deployment of troops) until Germany actually entered their territory.
24 Terraine, op. cit., 146.
26 Lt-Gen Alan Brooke, then commanding the II Corps of the British Expeditionary Forces, had, after visiting French troops, ‘most unpleasant apprehensions as to the fighting quality of the French in this new war.’ Cited in: Horne, op. cit., 92.
27 Cited in: Horne, op. cit., 489. According Horne after the end of the campaign 90,000 French soldiers were killed, 27,000 German soldiers were dead; ibid., 509-10.
29 Horne, op. cit., 102. This was correspondent to the French style of leadership: remaining in the command posts, managing the battle and not personally leading the troops. See Doughty, op. cit., 329-31.
30 Subsequent, considerably more German commanders were killed in action than their French counterparts.
32 Bloch continued his verdict about the French leaders as follows. “[T]he German triumph was, essentially, a triumph of intellect—and”, written in 1940, he foresaw “it is that which makes it so particularly serious.” ibid., 36.
33 Van Creveld, op. cit., 164.
36 Smith, op. cit., 134. Furthermore, it was not the doctrine or the operational methods which decided about defeat or victory, it was the fact that the Allies started their military industrial machines, as Smith pointed out. Germany lacked of raw materials and petrol. “Barbarossa was no longer Blitzkrieg: it had become an old-style industrial battle.” Ibid., 135.
The Blame Game:  
An analysis of the outbreak of war in 1914

LT COL P A MAYNARD RM

Introduction

"Of all the cruelties that people have inflicted on one another, the most terrible have always been brought by the weak against the weak."1

Stood on the balcony of the Sarajevo Town Hall in 1936, novelist Rebecca West reflected on why events there on 28 June 1914 had led to the outbreak of a world war. "I shall never be able to understand how it happened." 2 On the implicit point within West's observation, Christopher Clark elaborates, "it was not... that there were too few facts available, but that there were too many." 3 If this was true in 1936 it is even more the case today.

Few events in history have been more thoroughly researched, analysed and debated than the First World War. It is a densely populated historiographical terrain, one interpreted in almost every discernible way. The scale of the literature is extensive, with tens of thousands of books published in the English language alone. The breadth and depth of scholarly interest has reflected the war's status as "Die Ur-Katastrophe" 4 - the original catastrophe. It resulted in over thirty million casualties, 5 the collapse of four empires, and accelerated the demise of Europe's global pre-eminence. Fritz Stern argues it was "the first calamity... from which all other calamities sprang." 6 It brought about a complete unhinging of the international system with extremely toxic and long-lasting consequences, some of which are still being felt today. It is, therefore, not surprising that so many have sought to understand and explain the cause of such a cataclysm and, in many cases, have sought to apportion blame. What is remarkable, after almost a century, is the continuing lack of scholarly consensus.

Margaret MacMillan suggests, "there are many, perhaps too many, explanations for why Europe chose war over peace in 1914." 7 Amongst others, John Langdon and Annika Mombauer have both published works that seek to synthesise the 'long debate' and explain its discourse. 8 It is a discourse that reveals much about the nature of history, about the evolution of historiography, about the contemporary context of an observer, and the political, social and philosophical influences that determine his or her particular interpretation of the evidence selected. As E. H. Carr argued, "the historian will get the kind of facts he wants. History means interpretation." 9 Using the words of George Gooch, Mombauer notes, "it is part of the tragedy of the world war that every belligerent can make out a case entirely convincing for itself." 10

Today the war itself has moved beyond living memory but, because the truth about the outbreak of war in 1914 remains contested and ambiguous, it is a study that retains a timeless resonance for those wishing to understand how the international system can collapse in catastrophic failure.

Relieved of the political and social sensitivities of earlier periods, historians today continue to seek new ways of interpreting the dramatic events that unfolded in the summer of 1914. Clarke argues that new political and strategic eras provide distinct vantage points through which to view the past. 11 He highlights how the Yugoslav wars in the 1990s provided a reminder of the lethality of Balkan nationalism, and how the terrorist attack on the United States in September 2001 provided an opportunity to consider the power of single events to change global politics. MacMillan highlights the centenary echoes of rapid globalization, the competition for resources, economic crisis, shifting power balances, and the assumption that large-scale war is unthinkable. 12 Whilst MacMillan’s argument does suggest an element
of ‘presentism’, the point for Clarke is that the contemporary world enables an observer to see the past in unique and possibly clearer terms. The closer one is able to view the events of 1914 through the perspective of those whose decisions in 1914 determined the fate of the world, the better the assessment will be. This requires an approach that explores the contingent nature and interaction of proximate factors, supported by an examination of the effect from broader, underlying and deeper forces.

This paper analyses the proximate origins of the war that broke out in July 1914, the deeper contextual forces that set the necessary conditions, and the subsequent debate that has concealed the truth. It determines why a regional conflict in the Balkans evolved so rapidly into a world war, and whether or not war could have been prevented.

The paper argues there was neither anything inevitable about the outbreak of war, nor did the nations slide inadvertently into it. All the belligerents were responsible to a greater or lesser degree, all wilfully risked war, and all retained alternative choices up until the fighting broke out. The paper concludes that the real guilt lies with the elite rulers; the civilian and military decision-makers who recklessly chose war over peace in July 1914. It makes the case that they accepted war in order to stave off the forces of change, both within and without, which threatened their existing order, power and authority.

The first section, Descent into the Abyss, explores the July Crisis that precipitated war in 1914. Using a mix of older and more recent sources, the various actions and decisions of the belligerent nations are analysed chronologically. Gordon Martel suggests the choice about where one starts an analysis of the outbreak of the First World War “reveal[s] something of [their] own assumptions and prejudices.” This analysis begins in the Balkans, not just because that is where the countdown to war began, but because this approach supports the case that war was not inevitable, that Sarajevo was not just a simple pretext. The section argues that the July Crisis was complex and fast moving. Although the account is necessarily brief, it demonstrates the intense and multi-faceted challenges faced by the belligerent nations. The outcome was the result of decisions based on highly contingent factors, and made by elite coteries who held significant power.

The second section, Unspoken Assumptions, analyses the underlying context within which European rulers and statesmen made their decisions for war. This section underscores the importance of a deeper contextual understanding, identifying the conditions that help explain the increase in tensions between the main belligerents, and the development of attitudes that may have helped determine the path to war. It identifies, and then considers, some key political, diplomatic, sociological, philosophical and military factors.

The final section builds on the previous analysis in the paper, examining the historiography and utility of The Blame Game. The analysis considers the imposition of war guilt in 1919, the revisionist consensus that emerged in the interwar years, and the impact of the Fischer controversy in the 1960s. Throughout, it highlights how contemporary political issues have influenced the debate. Analysis of the latest discourse on the war’s origins highlights how fractured the Fischer consensus has become, and how the 1930s orthodox view has been partially revived. Final reflections focus on the underlying theme throughout the paper, which casts the various belligerent leaderships into the spotlight.

**Descent into the abyss**

"The public… waited on the sidelines as a handful of men in each of Europe’s main capitals juggled with fateful decisions."

It was Otto von Bismarck, Germany’s first Chancellor, who famously remarked that the next great European war would be triggered by “some damn fool thing in the Balkans.” Bismarck’s insightfulness was later proven when the outbreak of the First World War was precipitated by the infamous assassination of Archduke Franz Ferdinand, Habsburg heir to the Austro-Hungarian throne. The subsequent thirty-seven day crisis engulfed rulers and policy-makers across Europe. The Chancellor’s remark can also be used to reinforce a view that there was something inevitable about the outbreak of a
war in Europe, and that any proximate cause provided little more than a ‘tinderbox’. Explanatory theories that stress more deterministic influences, the imperialist ambitions of one or more of the Great Powers for example, tend to treat the terrorist attack in Sarajevo on 28 June 1914 as “a mere happenstance pretext.”¹⁷ Such theories are too simplistic. There was nothing inevitable about war at the end of June, or even through much of July in 1914. As Mombauer argues, “Right until the last moment, some were desperately trying to avoid the outbreak of war and to resolve the crisis. That war finally broke out was less the product of fate or bad fortune than the result of intention.”¹⁸ A survey of the diplomatic prologue to the war, the ‘July Crisis’, provides ample evidence in support.

Assassinations were relatively commonplace at the time. In the previous fifteen years there had been at least forty successful political assassinations, including seven Heads of State and Presidents. In 1914 the assassinations of socialist leader Jean Jaurès and Le Figaro editor Gaston Calmette were arguably both more shocking than the Archduke’s death in Bosnia. Calmette was front-page news in July 1914, whereas the death of the Archduke made little immediate impact beyond central-south Europe. Neiberg highlights how the Irish Times reported the Sarajevo assassination as “merely another tale of blood in the annals of the ill-fated House of Hapsburg.”¹⁹ Franz Ferdinand’s assassin, Bosnian teenager Gavrilo Princip, was one of seven assassins who lined the streets alongside the river Miljacka on that fateful Sunday. They were enthusiastic amateurs and only fluked success following an unpredictable navigation error by the Archduke’s driver. The group had been trained and supplied by a secret Serbian terrorist organisation, known as the Black Hand, a group who desired pan-Slav unification. The leader of the Black Hand and also Serbia’s Chief of Military Intelligence, Dragutin Dimitrijević, believed that the killing of royalty could bring about political change. Such tactics had worked for Serbia before, in 1903.²⁰ Furthermore, the Black Hand was only one of the radical groups operating in the Balkans at the time; the Internal Macedonian Revolutionary Organisation was also flourishing.²¹ Political terrorism was therefore neither new nor particularly unique in 1914. It was also not the first crisis of the period in the Balkans. The Great Powers had diplomatically contained several previous crises, such as the Austro-Hungarian annexation of Bosnia-Herzegovina in 1908, and two Balkan wars in 1912 and 1913. So what was different this time?

The response of Austria-Hungary was crucial. Few grieved for the death of the Archduke, but Vienna was outraged by what was widely perceived to be a humiliating attack by Serbia. An investigation and an appropriate response to the attack would have been understandable, but Austria-Hungary saw an opportunity to solve what they saw as the Serb problem. As Hastings argues, “Austria made an almost immediate decision to respond… not because its leaders cared a fig for the… slain Archduke… but because [it] represented the best justification they would ever have for settling accounts with a mortally troublesome neighbour.”²² The Habsburg elites were motivated by fear, by “a deep suspicion that if the regicide at Sarajevo was not countered with strong military action, it, like the Ottoman Empire, would rush into irreversible decline.”²³ Ironically, it was not a view the Archduke himself would likely have held. Tragically, as MacMillan argues, “the assassination had removed the one person close to the emperor who might have counselled restraint in those last weeks of Europe’s long peace.”²⁴ Nobody in Vienna appeared interested in the fact that Pašić, the Serbian Prime Minister, seemingly opposed to the Black Hand movement,²⁵ had tried to prevent the Sarajevo attack by sending “a veiled warning to the Austrians… which apparently failed to register.”²⁶ Strachan points out this may have just been duplicitous Serb politics.²⁷

Although the hardliners, such as Hötzendorf, had gained the upper hand, it was recognised in Vienna that an attack against Serbia would almost certainly draw a hostile response from Russia. The Austrians were therefore reliant on support from Berlin. The German Kaiser, Wilhelm II, was upset by the loss of Ferdinand who had been a close friend, but he too recognised the opportunity in the tragedy to shore up the position of the Central Powers in Europe. In a moment of “uncharacteristic decisiveness”²⁸ he asserted: “The Serbs will have to be straightened out, and soon.”²⁹

Berghahn describes the July Crisis as an evolution between three phases: First, the development and initial implementation of a German-Austrian plan; second, the gradual unravelling of that plan; and third, the descent into war. Whilst there have been innumerable accounts of the crisis, this structure helps to expose the interplay of contingencies, the conscious risk taking, and wishful thinking that existed.
By July 1914, several influential members of Berlin’s ruling circle, not least the General Staff who reported directly to the Kaiser, were advocating decisive Austrian action in the Balkans, mainly because of the estimated force preparedness of both Russia and France. From Germany’s perspective, a combination of Austro-Hungarian decline, Italian unreliability, regeneration of Russian military power, and secret intelligence reports of British-Russian naval negotiations gave rise to a sense of encirclement and vulnerability. Policy makers in Berlin “did not actually want a European War [but] they were certainly willing to risk it.”

Confidence stemmed from the wave of international opinion in support of Austria, and the fact that Russia had been deterred in the Balkans before, when Germany made it clear it was supporting Austria-Hungary’s annexation of Bosnia-Herzegovina in 1908. Diary entries from the July Crisis, such as those of Plessen, Chief of the German Military Cabinet, demonstrate that Russian interference was judged unlikely.

Following a request for support from Franz Joseph, the Habsburg Emperor, the German Crown Council decided, “the dual monarchy’s action [in the Balkans] was its own affair, but it was assured of German support in the event of Russian intervention.” The infamous ‘Blank Cheque’ had been written. The issue, as Strachan highlights, was that the German decision “followed from previous events rather than from a projection as to the future.” The challenge for Germany in July 1914, and for Chancellor Bethmann-Hollweg in particular, was to ensure the underwritten risk could be managed, that the “leap into the dark”, could be made to work.

For both the Kaiser and Bethmann-Hollweg, swift action was required to prevent intervention by the other major powers and “long-winded diplomatic preliminaries.” The intent was to destroy Serbia quickly in a localised war, “to achieve a ‘fait accompli’ and then to be friendly towards the Entente.” Bethmann-Hollweg realised that international sympathy for Austria would dissipate quickly, and a drawn out response would appear much more suspicious to the Entente powers. A perception of calm was deliberately created, with many of the key decision-makers in both Berlin and Vienna, especially military chiefs, continuing with summer holiday plans. The approach was successful, at least, in preventing interference by outside powers until the decisive act: Austria’s ultimatum to Belgrade on 23 July 1914.

Austria’s preliminary move had taken longer than Germany had wished. Concerns about establishing a juridical basis for attack, especially from the likes of Tisza, the Ministers-President of Transleithania, delayed decisive action until the Austro-Hungarian Council of Ministers gave final approval on 19 July. An additional delay of four days was incurred by the need to ensure delivery of the ultimatum did not coincide with the visit to St Petersburg by French President Poincaré. It was a delay supported by Berlin but one that did little to prevent it being seen as premeditated. The nature and slowness of the Austrian response had fatal implications for Bethmann-Hollweg’s fait accompli plan, significantly compromising what had been high-risk preconditions for success. The cumulative effect helps explain why some historians later viewed the Austro-German approach as so belligerent:

“It may be that it was this virtual coincidence of the decision to wage war against Serbia at the conscious risk of a world war on 5 July and the practical inevitability of such a major war as a result of the procedure adopted [by Austria-Hungary]… which induced Fritz Fischer to assume that the Central Powers wanted a world war all along.”

With the intervention of the other major powers from 23 July, the “German house of cards was beginning to collapse.” The Austro-Hungarian approach, supported by Berlin, had led to an ultimatum that sought to make war unavoidable for Serbia. The British Foreign Secretary Lord Grey described it as “the most formidable document I had ever seen addressed by one state to another that was independent.” Somewhat unexpectedly, possibly buoyed by indications of support from Russia, the Serbs acquiesced to almost all the terms on 25 July. The only objection was to the demand for Austro-Hungarian officials to lead the Serbian enquiry into the Sarajevo assassination. Whilst sufficient to justify subsequent putative measures by Austria, the Serb response made such action “look suspicious in the eyes of those European powers who wanted to try to preserve the peace.” This is not to suggest that the Entente powers of France and Russia wanted to preserve peace. As McMeekin makes clear, there was as much collusion
between St Petersburg and Paris as there had been between Berlin and Vienna. It is just less documented.

Although frustrated that the situation was slipping out of control, the Kaiser had no appetite to accept another diplomatic defeat. Having failed to secure swift action by the Austrians, Bethmann-Hollweg's position was “too weak to beat retreat.” The only option appeared to be a continuation of the high-risk approach. Appeals were made to St Petersburg, Paris, and London in an attempt to localise the conflict. Bethmann-Hollweg told German ambassadors in the European capitals, “any intervention by another power will, in view of the divergent alliance commitments, lead to incalculable consequences.” But the cracks were starting to show. Convinced that the Central Powers aimed to annihilate Serbia, the Russian Council of Ministers ordered the ‘Period Preparatory to War’ on 24 July. In France troops were recalled from leave and withdrawn from Morocco. “What Bethmann-Hollweg had recognised as a potential danger and consciously accepted as a risk all along, had at last come about: it was impossible to localise the conflict.”

As the last week of July progressed, diplomatic control of events began to wane. Belgrade, having accurately predicted Vienna’s lack of appetite for a diplomatic solution, mobilised its army on 26 July. The Austro-Hungarians quickly followed suit the next day, although it was unlikely to be able to commence operations for at least two weeks. Berchtold was under pressure from Berlin to prevent the space for international mediation and he, therefore, declared war on Serbia twenty-four hours later. The third Balkan war had begun. In Berlin, Bethmann-Hollweg and the Foreign Ministry were under increasing pressure from the General Staffs to consider the military option. Diplomatic attempts to keep the peace quickly evolved into efforts to set favourable conditions for war. Bethmann-Hollweg recognised this meant a need to ensure German domestic unity and a securing of British neutrality. Bethmann-Hollweg’s first task was partially achieved through the Bethmann-Südekum accord, which secured the support of the German Socialist Party and trade unions. As far as the masses were concerned, they would only unite behind the Kaiser in a defensive war against an external aggressor. On 30 July Tsar Nicholas II ordered a general Russian mobilisation, giving German leadership the narrative it needed.

Like Bethmann-Hollweg’s plan for Austria-Hungary, the Tsar’s decision to mobilise was criminally high-risk. As McMeekin argues, the Tsar, alongside Sazonov, Sukhomlinov and Yanushkevitch “chose consciously to mobilise Russia’s colossal armies in full knowledge that they were risking war with Germany by doing so.” Russia was secretly preparing for a European war from at least 25 July. McMeekin also questions why mobilisation of both the Baltic and Black Sea Fleets were deemed necessary to confront the Austrian threat to Serbia. The chief of Russia’s military mobilisation, Doborolskii, later recalled of the period; “the war was already a settled matter, and the whole flood of telegrams between the governments of Germany and Russia represented merely the stage setting of a historical drama.”

But did mobilisation really equate to war? It certainly had a decisive effect on the German public. The entire German press reflected “the seemingly inevitable war… [as] a defensive crusade against the dark forces of Slavdom and oppressive Tsarist autocracy.” McMeekin accuses Russia’s political leaders of playing a game of deception during the last days of July, specifically deceiving London about Russia’s less than peaceful intentions. Unlike Berlin’s failed plan to achieve a quick and localised fait accompli in Serbia, McMeekin accuses Russia of deliberately seeking a European war and of manipulating Britain, as a powerful ally, in to it. There was a single brief moment when the Tsar considered halting Russia’s mobilisation, a result of the ‘Willy-Nicky’ telegrams in the final days of the crisis, but the correspondence between ruling cousins could not keep pace with the fast-moving events. When the Tsar revealed that Russian preparations had begun five days earlier, an error of admission also made by Foreign Minister Sazonov to Portales, the German Ambassador to Russia, all doubt had been removed in Berlin about the likelihood of war. Importantly, Germany’s decision-making elites now realised that Russia had stolen the march.
“Mobilisation for Germany did mean war.” Their operational strategy, devised by the Army Chief of Staff Count von Schlieffen in 1905, and later modified by Von Moltke, provided for a two-front war against Russia and France - on the assumption that the Entente defensive alliance would be honoured. In short, the plan amounted to a swift advance to secure a quick capitulation by France. This would enable Germany to subsequently mass a larger force against the Russians, who would take much longer to mobilise in the East. This single option presented to the Kaiser by the German Army Chiefs in 1914 required an invasion of both Luxembourg and Belgium and, therefore, a violation of the 1839 Treaty of London. Its lack of flexibility and subsequent execution would provide a final German blunder in its attempts to minimise the scale of the conflict, hammering a final nail in Bethmann-Hollweg’s plans to secure the neutrality of the British.

The British Foreign Secretary, Lord Grey, had maintained a cautious approach to the July crisis. He had “suggested that the issue could be resolved at the conference table, but [his] mediation proposals and attempts to preserve the peace were not taken up by Vienna or Berlin.” As the crisis culminated towards the end of July, the British position on the impending war increased in importance for the other major powers. While German and French Generals may have been indifferent to the threat posed by the six Divisions of the British Expeditionary Force, Kaiser Wilhelm, Bethmann-Hollweg, Tsar Nicholas II, Sazonov, and Poincaré all recognised the pivotal impact the Royal Navy could make on the war. The British cabinet were split between the peace party and the interventionists. Domestic pressures, such as the Irish Home rule question, weighed heavily on the government, and the 1912 Anglo-French Naval agreement was already threatening to suck a reluctant nation into war.

King George V had reassured the Kaiser, via Wilhelm’s brother Prince Heinrich, that Britain would remain neutral. But Bethmann-Hollweg understood the British parliamentary system better and made a final plea to the British government on 29 July. His offer included an undertaking by Germany not to annex any French territory, but in doing so he also spelt out the German war plan and the intent to march through Belgium. This demonstrated “how far a cool appreciation of the harsh realities had once again become displaced by wishful thinking.”

It is debatable how far Britain’s reluctance to declare its position until the end of the crisis helped to fuel it. Mombauer argues, “Grey’s policies should not be seen as a cause of war. After all, this hesitant attitude was motivated by the desire to avoid an escalation of the crisis.” Since British territory was not being directly threatened by the crisis, any decision to intervene had to be debated by more people than was the case in the other European capitals. Despite clear threats to grand strategic imperatives, unifying the British government and people behind a case for intervention was challenging. The German declaration of war against Belgium on 1 August made the job much easier. As Strachan argues, the “commitment [to Belgian neutrality] became the bridge which allowed Realpolitik and Liberalism to join forces.” Although more nations had still yet to become embroiled, by 4 August 1914 the world was at war.

This brief survey of the crisis, which ultimately caused the First World War, highlights the complex, multifaceted and fast moving nature of events. “Events had moved faster than the imagination of the politicians.” But this short summary necessarily belies the true depth of complexity that existed. Joll and Martel point out that Luigi Albertini’s ‘The Origins of the War of 1914’ devotes some 1,400 pages to the period between 28 June and 12 August. There are thousands of diplomatic documents and personal accounts that depict different perspectives and aspects of the story. Following decades of academic scholarship, they have each been poured over, interpreted with hindsight, and judged with varying emphasis and importance. Notwithstanding, it is evident that the July Crisis was an intensely interactive, multi-agent challenge. The outcome was the cumulative result of decisions taken by small groups of rulers and policy-makers, each of which held a very large share of power. As Strachan argues, “what remains striking about those hot July weeks is the role, not of collective forces nor of long-range factors, but of the individual.” It is because of the tragic consequences following the decisions of such individuals that many have sought to explain them.

In 1923 Winston Churchill reflected, “the spring and summer of 1914 were marked in Europe by an exceptional tranquillity.” Why peace evaporated so rapidly is just one of the many questions that
emerge from analysis of the July Crisis. Indeed, whilst the crisis created its own momentum, determined by a chain of iterative and interactive decisions, such decisions did not stand in isolation. Why did Austria-Hungary blame Serbia so quickly for the Sarajevo attack and seek such ruthless justice? Why did Germany feel the need to provide unconditional support, wilfully risking war? Why did Russia risk war by mobilising before anyone else, and why did France support it? Why was insufficient effort made to mediate and or consider compromise? Such questions cannot be substantially addressed by looking solely at the events of July 1914.

Joll highlights how many of the men at the forefront of the July Crisis felt they had been “carried away by the tide of history.”62 Such fatalism may represent a weak and unsatisfactory explanation, but it does suggest something about the way such individuals felt at the time, certainly at the end. In this respect, it is important that the events of the July Crisis are placed into a broader and deeper context.

Unspoken assumptions

“Again and again in the July crisis one is confronted with men who suddenly feel themselves trapped, caught up in a fate they are unable to control.”63

In his celebrated 1968 inaugural lecture for the Stevenson Chair, titled ‘1914: Unspoken Assumptions’, James Joll concluded, “it is only by studying the minds of men that we shall understand the causes of anything.”64 He argued that the behaviour of those who brought about war in July 1914 could only be fully understood if one explored the ‘assumptions’ on which their decisions were based. He was encouraging a deeper understanding of events, a multi-disciplinary approach to international history. It is a theme reflected by the current Stevenson Professor, David Stevenson, in his own inaugural; “International history cannot be understood without reference to national history – although the converse also applies.”65

Many studies into the causes of the First World War have adopted such an approach, seeking to understand the logic (and illogic) of the course of events. As early as 1928, Sidney Bradshaw Fay suggested five underlying causes: “the system of secret alliances, militarism, nationalism, economic imperialism, and the newspaper press.”66 Other than the last one, such “Doomsday Machine”67 factors have endured in the war origins scholarship. There are others too, which focus on the influence of intellectual ideas or social conflict. All are contentious but add value to the debate. An expansion of analysis, beyond the proximate and contingent causes of the war to longer-term and more deterministic origins, risks distortion and over-simplification of the reality. The best one can hope to identify are conditions that help to explain an increase in tension, or the development of attitudes that may have helped determine the path to war.

This section explores the underlying and deeper forces that conditioned and shaped the course of the July Crisis. It will consider some of the political, diplomatic, sociological, philosophical, and military factors. The value of such analysis does not reside in exposing the relative merit of any particular cause or theme, but in the way the contextual factors interact, interconnect, and converge. As Lebow points out, “timing was everything in 1914.”68

In the preceding hundred years, only the Crimean War had interrupted the peace between Europe’s Great Powers.69 The Concert of Europe had successfully prevented a recurrence of the catastrophe that beset Europe during the Napoleonic Wars. When wars did occur, such as those for imperial interests or those that conditioned the unification of Germany or Italy, they were fought for clear political goals and limited in nature. Clausewitzian logic prevailed. However, by the turn of the century the Concert of Europe, as the predominant regulator of world affairs, was being undermined by both globalisation and the rise of Germany. The interrelated impact of both helps explain the increase in international tension before 1914.

By the first decade of the new century the international system had expanded in profound ways. The second industrial revolution had delivered new technologies that transformed international transport
and communications. Economies had become increasingly interlocked and interdependent with unprecedented movement of capital and trade between states. And international institutions, such as the Hague Conferences, attracted expanding state representation. As Clark highlights, “the ushering in of the twentieth century marked the beginning of globalized international relations.” But the litmus test for great powers remained their ability to pursue their own will. A significant impact of globalization was the culmination of imperialism and the emergence of new non-European powers. In the Far East new political forces were gaining influence and power. Japan had defeated both China and Russia in 1895 and 1905 respectively, and the United States had annexed the Philippines, Guam and Hawaii during the 1898 Spanish-American War, before asserting an ‘Open Door’ policy for China. Importantly for Britain, Russia’s turn towards the Pacific effectively ended the Great Game in Asia, a reality later confirmed by the 1907 Anglo-Russian convention. Such developments led to new ideas about a global balance of power, which subordinated the traditional conception of a “struggle for mastery in Europe.” As both Britain and Russia pursued new global strategies, which necessitated new alliances and accommodations, Europe became increasingly unstable.

If imperial rivalry characterised the years at the turn of the century, there remained a basic disequilibrium at the European core. This was largely the result of Germany’s unification in 1871 and of the imbalance generated by rising German power in central Europe. The creation of the Franco-Russian entente was the major geopolitical response. Germany shared the view that international politics was a competition between world powers but, with much of the globe already carved up, realised it was playing catch up and that time was running out to secure its place in the sun. “While Britain came to preside over half the globe, and both Russia and the United States relentlessly filled out their continental hinterlands, Germany was expected to remain locked within the tight frame of Europe’s traditional balance of power.” Germany’s embrace of Chancellor Bülow’s Weltmachtpolitik was simply a response to such a predicament. As Clarke argues, “Germany must be viewed as reacting to a new condition of the balance of power which had already emerged, rather than as initiating it.”

As the Concert of Europe gave way to an emerging era of competitive global politics, Bismarck’s intricate alliance system, designed to safeguard German unification, was evolved by his successors to align the European powers into two opposing blocs. According to Stevenson, the Concert “could work only when all the Powers wanted it to, [but] once they were polarised into two alignments it was crippled.” In particular, the German decision not to renew its ‘reinsurance treaty’ with Russia was “a terrible mistake.” The system became hinged precariously on Franco-German antagonism, competing interests of Austria-Hungary and Russia in the Balkans, and German fears of being encircled. Some would add Russian fears of being enveloped too. While the so-called “diplomatic doomsday mechanism” has often been cited as a key factor in the cause of the First World War, its importance should not be overstated. As Herwig points out, “none of the decisions for war [were] mandated by treaty obligations.” Indeed, in July 1914 some parties elected to offer far greater support than their commitments specified. Almost all mobilisations were motivated by “brutal self-interest” as opposed to an altruistic honouring of commitments. It is therefore more useful to view Germany’s support to Austria-Hungary in July 1914 in terms of its own sense of isolation and encirclement, grasping to hold on to their last reliable partner. And Russia’s backing of Serbia should be viewed in terms of protecting the final link to vital strategic interests in the near East. Notwithstanding, the dynamics of the alliance system would play a fatal role in determining the military responses to the July Crisis, a point that will be addressed later.

A key issue that undermines the alliance system argument is why so many crises had been successfully resolved, or at least contained, in the years before 1914. By analysing these crises, and the cumulative effect they had on the perceptions of the European powers, a better understanding of the July Crisis emerges.

The pressures of the emerging global era were manifesting themselves in several ways. Perhaps the most significant was the realisation by Britain that it could no longer remain in ‘splendid isolation’. “Its [1902] treaty with Japan was a way of testing the waters, to see if it wanted to plunge further into the entanglements of alliances.” The need to avoid being pulled into the Russian-Japanese war would force Britain’s hand. An understanding with France in 1904, alongside the later 1907 convention with Russia,
effectively pulled Britain towards a Triple Entente. Britain’s accommodations with France and Russia sought to protect imperial interests. From Germany’s perspective, the developments simply reinforced its own sense of isolation.

It is in this context that we can partially understand Germany’s seemingly aimless, even reckless foreign policy in the years before 1914. Early blunders, such as the Krüger telegram, had already begun to alienate the British, but it was in Morocco that Germany inadvertently helped to cement the Entente Cordiale. Exploiting Russia’s defeat by Japan, Kaiser Wilhelm II failed in his attempt to impose German influence at Tangiers in 1905, the intended effect being to undermine French trust in both Russia and Britain. Britain intervened, and did so again six years later when an increasingly assertive Germany sent a Warship to Morocco. The Agadir deployment ostensibly sought to protect German interests during the controversial French intervention at Fez, but the British perceived it as an act of naval provocation. Responding to the crisis, David Lloyd-George warned, “if Britain were to be forced to choose between peace on the one hand and the surrender of her international pre-eminence on the other, then I say emphatically that peace at that price would be a humiliation intolerable for a great nation like ours to endure.”81 Both British antagonism with Germany and a deepening bond with France were forming. France was learning that Germany only understood the language of force. And in Germany, not only was frustration growing over a succession of diplomatic defeats, the Kaiser, as Clarke highlights, had “disassociated himself from his chancellor and re-emerged as a threat to the policy-making process.”82

Perhaps the most significant consequence of the 1911 Morocco Crisis was the domino effect it would have across a gradually dissolving Ottoman Empire. More than any other factor, it is this that links the pre-war crises to the July Crisis. Inspired by French activity in Morocco, “it was the Italian attack on Libya in 1911 that flashed the green light for the all-out Balkan assault on the Ottoman periphery.”83 Clarke recalls the words of Spalajkovic, political head of the Serbian Foreign Ministry, “all subsequent events are nothing more than the evolution of that first aggression.”84 It could even be argued that the demise of Ottoman power was a more significant threat to European equilibrium than the rise of German power, and that the First World War “could very easily be labelled the War of the Ottoman Succession.”85 That first aggression by Italy exhausted the Ottoman state, and left the Balkans wide open.

Austria-Hungary had already exploited Ottoman retreat to formally annex Bosnia-Herzegovina in 1908. German backing of this move again exploited a war-weary Russia. It was a humiliation that threatened pan-Slav solidarity, one Russia would prevent from happening again by launching a military reconstruction programme. Russia’s reawakened focus on the near East, and Constantinople (Tsargrad) in particular, led to support for the Balkan League in its war and subsequent victory over the Ottomans in 1912. The coalition of Greece, Bulgaria, Serbia, and Macedonia quickly turned on each other in a second war over disputed territorial gains. A number of important consequences emerged: A much expanded and aggressive Serbia now posed a greater threat to Austria-Hungary; the Bosporus Straits closure in 1912 had started a navy arms race in the Black Sea; Russia had become suspicious of German military support to Turkey; Germany realised that the position of Austria-Hungary was weakening; and Britain had become convinced of its mediation credentials. Above all else, due to continued tension, ethnic persecution and huge numbers of both Christian and Muslim refugees, another Balkans war was widely anticipated. Most expected a war pitting Greece against Turkey. What Europe got was a dead Archduke on 28 June 1914.

Some of the unspoken assumptions that pervaded the diplomatic crisis in July 1914 were being formed in the crises of the preceding decade. It is true that these crises were contained and largely resolved through diplomatic means, but this misses the point. Referring back to Joll, “when political leaders are faced with the necessity of taking decisions the outcome of which they cannot foresee, in crises which they do not wholly understand, they fall back on their own instinctive reactions.”86 One might suggest that leaders’ instincts in July 1914 reflected the different experiences they had drawn from the previous clashes. An appetite for mediation had reduced in favour of bold posturing. Stevenson argues, “The [Balkan War] had demonstrated to [Sazonov] that he could back up diplomacy by heightening military preparedness, but to the Austrians that even armed diplomacy against Serbia did not work, and to both the Austrians and the Germans that the conventional diplomatic device of mediation via Great Power
conference did not operate in their interests.”87 Poincaré illustrates the point when, in July 1914, he made
the assumption Germany’s “prodding [of] Austria-Hungary… was to carry on the policy of the second
Moroccan crisis and split the Entente.”88 Britain’s diplomatic assumptions were similarly exposed. With
London balanced precariously between isolation and alliance commitment, July 1914 proved that
neutrality no longer attracted peace-inducing powers. Clarke provides the fundamental point: “It was the
changed context in which the alliances operated that was to be so decisive, and destructive, in the
summer of 1914.”89

So far the analysis has revealed a broader political context to help explain the outbreak of war in 1914.
While globalization altered the international order, the impact of rising and declining powers destabilised
it. The diplomatic precedents to July 1914 are clearly traceable and help explain the increase in
international tension, but such analysis provides only one perspective on what was influencing the minds
of those that took the world to war. There are other assumptions to be explored. For example, why did
all states in the July Crisis explain their intervention on the basis of self-defence? Why was the recourse to
war so rapid and attempts to prevent it so ineffective? Most importantly, why were the people of Europe,
having enjoyed relative peace for so long, so willing to respond to the call to arms? To answer these
questions, one must delve deeper than the political presumptions, into the climate of social issues and
the influence of intellectual ideas. As Tuchman argued, “the diplomatic origins… are only the fever chart
of the patient; they do not tell us what caused the fever. To probe for underlying causes and deeper
forces one must operate within the framework of a whole society and try to discover what moved the
people in it.”90

This paper has already highlighted the impact of globalization on international relations, but the
industrial and scientific revolutions that underpinned it affected much more. States had to contend with
the domestic consequences of rapid growth and change. These changes brought social pressures in the
form of population growth, urbanisation, increasing poverty, class inequality, and secularisation. But
they also induced mass society, with a growth in education, literacy, popular press, access to consumer
goods, and spectator sport. Increasingly citizens began to identify themselves with their nation, a new
larger community in which they had a stake. As society expanded, so too did participation in the political
franchise. Across Europe new political groups and movements began to press for greater rights, a direct
threat to the established aristocratic orders. For some, such as Norman Angell in his 1910 book ‘The
Great Illusion’, the age of capitalism and interconnected markets meant war could no longer hold a
rational or useful purpose.91 Angell was not alone; Stefan Zweig, a middle-class Austrian writer, referred
to the period as “the Golden Age of Security.”92 But Tuchman exposes the truth: “A phenomenon of such
extended malignance as the Great War does not come out of a Golden Age.”93 The deep unease and
underlying tension of the period was widespread.

The demands of social and national advancement forced the state to assume greater social control, a
move that led to the nationalisation of societies, but also a challenge to the authority of the traditional
ruling elites by the new industrialists. The challenge came primarily in the rise of socialism. By 1912 the
socialist party in Germany, the SPD, “constituted the largest single party in the Reichstag.”94 The French
socialists were gaining seats and the British Labour party was also growing in influence. By the July Crisis
there were 4.2 million socialist members worldwide, united in the Second International.95 The prospect of
mass revolt against so-called ‘imperialist war’ threatened to denude rulers of the military resource
required for war. The political challenge wasn’t just a socialist one either; syndicalists, anarchists,
suffragettes, and terrorists all forced the focus of political leaders on domestic affairs. For Europe’s old
aristocratic elites, “the whole interrelated but untenable system of power and privilege was moving
toward crisis and probably breakdown.”96 They discovered that nationalism provided an effective
antidote. As Gordon argues, “the more the forces of change pressed for reforms, the more the forces of
order took fright. Hence the alarming wave of fanatical chauvinism, mass demagogoy, and crude racism
that enveloped… from the right in the years before the war.”97

In the event, nationalism proved a more powerful force in the minds of socialist workers than proletarian
internationalism. Despite the best intentions of the Second International to unite against a European
war, socialist solidarity failed to materialise in July 1914. This can partly be explained by a lack of
preparedness. Ignorant of political secret diplomacy, and deluded by the effectiveness of their influence over previous crises, the speed of events left them as spectators. As a defensive war became seemingly unavoidable, many socialists reconciled themselves to the belief that “the nation itself could become a better society, and that the war might be the means to achieve that.”

Since the German SDP was the most influential socialist party, its decision to vote for war credits on 4 August 1914 sealed the fate of the Second International. As war broke out nationalists and socialists converged in political truce, whether in the Burgfriedenspolitik or L’union sacrée. Governments viewed war, at least in part, as an escape from internal challenges. Even in Britain, Prime Minister Asquith muted, “it is the most dangerous situation of the last fifty years. It may incidentally have the effect of throwing into the background the lurid pictures of civil war in Ulster.”

On its own, the collapse of socialism in 1914 only helps to explain why war was not prevented, not why it broke out. In the response of ruling elites to the changing social orders, one can detect a very clear attempt to shore up state power, through both increased jingoism and concomitant rejection of globalization. Elites increasingly feared losing power abroad as it would weaken their power base at home. As Clarke argues, “the need for domestic accommodations reduced the scope for international concessions on the key issues affecting the external standing of these states.”

But if all the Great Powers faced similar domestic challenges why did their foreign policies vary? In short, the answer lies in the differing political structures. Gordon explored this issue with reference to Britain and Germany, arguing “domestic conflict had a contrary impact in the two nations.” He explains that Britain’s internal turmoil, including growing attacks on the Patricians and a fiercely divided cabinet, led to a more moderate, underhand and defensive foreign policy. This explains Grey’s cautious and inconsistent approach to the July Crisis. As in all the states, a sense of decadence and unease existed in Britain, but the parliamentary system made it difficult for an extreme right influence to take root. Industrialisation over a longer period had absorbed the middle classes into a liberal state, there were few peasants available for political mobilization, and a tradition of anti-modernism existed. This was not the case in Germany. Rapid industrialisation had undermined an out-of-date state apparatus and provided the conditions for extremists on the right to exploit millions of outcast German peasants and a militaristic middle-class. It was the increasing domestic tension and challenges to the status quo, exacerbated by weak national identity, which caused the Kaiser to pursue the social imperialist solutions of Weltpolitik. Gordon suggests the German gamble with war in 1914 can best be understood as a response to years of frustrated attempts to temper domestic instability. He argues, “with one sudden desperate charge, the Reich’s leadership hoped to achieve that ultimate breakthrough in Weltpolitik that would safeguard the nation’s expansionist future and its conservative order at home.”

In explaining the ripe conditions for war in 1914, the political impact of social conflict only goes so far. It fails to explain why war was seen as the best solution. After all, for all the domestic problems, the impact of war would eventually prove to be far greater. It also fails to provide a satisfactory explanation for why societies rallied behind their self-obsessed leaders. To address these issues one needs to expose unspoken assumptions that existed within societies themselves. Ideas inspired by intellectuals, nurtured by nationalists, and given form by poets, playwrights, artists, composers, and the press. As Tuchman argues, “the Grosse Politik approach… is misleading because it allows us to rest on the easy allusion that it is ‘they’, the naughty statesmen, who are always responsible for war while ‘we’, the innocent people, are merely led. That impression is a mistake.”

Industrialisation not only led to social but also moral and intellectual change. “Science gave man new welfare and new horizons while it took away belief in God and certainty in a scheme of things he knew.” New ideas assumed powerful and pervading influence. For example, “Darwinian ideas had a great influence on the ideology of Imperialism.” Distorted interpretations of the evolutionary theory led to a view that states themselves were involved in a natural competition where only the fittest would survive. In an age of rapid progress and expansion, a belief emerged in a hierarchy of states, a belief in advancement at the cost to others, and of the need for superior martial qualities. Such views had been
nurtured in all the leaders of 1914 during their formative years, and help to explain the widespread
growth of militarism within societies. It manifested across Europe in a high regard for military virtues, the
explosion of military pomp and ceremony, and huge expenditures in military capability. Among elite
circles and extreme nationalists, war came to be viewed as a necessary and inevitable way to realise a
stronger and better world. Others took a less optimistic view, pointing to the detrimental eugenic effects
of war. Regardless, a general climate of war had become self-fulfilling in the minds of many during the
years preceding 1914. Joll illustrates this point by highlighting the fatalistic views of Kurt Riezler,
personal assistant to Bethmann-Hollweg:

“Eternal and absolute enmity is fundamentally inherent in relations between peoples; and the
hostility which we observe everywhere… is not the result of a perversion of human nature but of the
essence of the world and the source of life itself. It is not accidental, temporary and removable, but a
necessity… which will claim its place as long as there are men and nations.”106

Another general response to the unease of the fast changing era was a “subordination of science to art,
the triumph of Romanticism over the Enlightenment.”107 Underneath the façade of order and reason,
many people across Europe worried that progress “was leading to a coarser, more selfish and vulgar
society.”108 What emerged was a “fascination with the irrational, with emotions, with the supernatural.”109
Amidst the unease, war was welcomed as a desirable release from decadence. Thomas Mann viewed war as
“both a purging and a liberation.”110 Brooke described British recruits as, "swimmers with cleanliness
leaping, glad from a world grown old and cold and weary."111 Although misinterpreted and
misrepresented, the works of Friedrich Nietzsche were particularly influential, not least on the avant-
garde. The movement responded to perceived social decay with a sense of apocalyptic destiny. It could
be found in Italian Futurism, German Expressionism, and British Vorticism. It could be heard in the tone
poems of Richard Strauss, and seen in the viciousness of Wedekind’s plays. For radicals and romantics,
Nietzsche’s partially understood ideas “about politics and art, literature and life were inextricably
confused and contributed to a passionate sense of excitement and commitment.”112 The influence of
Nietzsche, including such metaphors as the Übermensch (Superman), helped to inspire thousands of
German and Russian soldiers, many of who carried personal copies of Also Sprach Zarathustra.113 Even
Gavrilo Princip was known to enjoy reciting the poem Ecce Homo.114 In July 1914, when the anticipated
opportunity finally came, people everywhere sought, “to transcend their ordinary lives and to find a new
set of values in what they believed would be a new and enriching experience.”115 This was not a class
phenomenon: “The working class went to war willingly, even eagerly, like the middle class, like the upper
class, like the species.”116

The general psychosis depicted above is important in explaining the conditions within which war was
able to break out, but it was a psychosis manipulated by elites. Nationalist groups embraced half-
understood philosophical ideas in their efforts to galvanise patriotic sentiment. As Strachan points out,
“Nietzsche himself was a self-confessed European, scathing about the nationalist preoccupations of
Bismarckian Germany.”117 Inconvenient truths mattered little as “popular ideals of patriotic duty
coalesced with those of the foreign-policy elite.”118 Through a forging of constructed history and
philosophy, nations were aligned against the culture and values of others. The historical creations of
Treitschke, for example, played a decisive role in rooting Germany as a nation, and were seized upon by
patriotic leagues.119

The ideological clash emerged most prominently in the growing divide between German Kultur and
British and French civilisation. The Germans came to view Britain as a nation in decline; a result they felt
had come about because matters of the spirit had been neglected. Individualism, materialism and
economic self-interest had subordinated the role of the state and had led to decadence. For Britain and
France, the German challenge was one against liberty, democracy, and freedom without regulation - the
natural law of Europe. Conversely, the Germans sought to preserve what they deemed to be the new
order of Western Europe; the German national spirit, Kultur, a rejection of capitalism. For them, freedom
was possible only by a subordination of the individual to the state.
At the heart of German Kultur, as Ritter argues, was romanticism. This is what unified the people in defence of the German nation, regardless of their hopes for what war might achieve. In this respect, war wasn’t seen as a way to escape domestic challenges, it was an attempt to solve them. The combined strengths of modernism and traditionalism would forge “an unprecedented increase in the vitality of the German organism.” Ritter further contends that Nietzschean influence and harsh militarism was not real Kultur, but forced on Germany by isolation and encirclement caused by the other Great Powers. In these terms, the war had been framed as a contest, not for principalities but for principles, for ideas. It is for this reason that the influence of the intellectuals cannot be ignored. As Strachan points out, “philosophers sought to make politics moral, but instead politicized morality.” By implication, the traditional limits of war were effectively removed. Evans explains why this meant there could be no quick or easy victory:

“War had been transformed in its aims from a Clausewitzian instrument of politics by other means, as it had still been in the hands of Bismarck, to an all-out conflict in which the aim was the total destruction of the enemy’s political system, way of life and even in some cases very existence. Ideology became paramount in warfare, where it had been almost entirely absent since 1815.”

The spread and manipulation of intellectual ideas in the pre-1914 era may have galvanised and oriented nations for war, but unprecedented levels of nationalism across Europe was more an immediate response to the outbreak of war than a cause of it. The impact of ideas, whilst significant, can be exaggerated. Among others, Neiberg argues there is no evidence to support a case of popular pressure driving leaders to decide on war in July 1914. Claims to the contrary were all ex post facto. In the last days of peace, Strachan argues, “crowds formed in the streets, squares and boulevards of large cities, less because they were enthusiastic for war and more because they were desperate for news…. In the rural areas the news did not travel fast, and in any case could seem scarcely relevant.” Given the small coteries in which decisions for war occurred, it is no surprise that the outbreak of war came as a shock to most of the population. Irrespective, the preceding analysis is necessary in order to understand why the forces for peace failed, why people everywhere responded willingly when the call to arms came, and why so few questioned the necessity or rationality of their respective national cause.

Deliberately following the political and sociological analysis above, the final body of assumptions to be explored are those about the military and war itself. Having considered the increasingly rigid dynamics of the alliance system, the impact of rapid industrialisation, and the influence of Darwin on national psyche, the rise of militarism marks the point where all convergence. It is this juncture in the analysis that can help to bridge some of underlying causes with the course of events in July 1914. Not least, it helps to explain why a regional conflict in the Balkans quickly transpired into a world conflagration.

“The situation [in Europe] is extraordinary… it is militarism run stark mad.” This American perspective, reported to President Wilson in May 1914, illustrates the extreme social pinnacle on which militaries across Europe had been placed. Nurtured by governments and nationalist pressure groups, such as the Army and Navy Leagues, the phenomenon was widespread in the pre-war years. Rulers embodied the trends, wearing uniforms in public, and German middle classes sought to imitate the practices of Prussian officer aristocracies, embracing duelling rituals and the pride of ‘Mensur scars’. Such was the blind respect for military officers, a German shoemaker, impersonating an officer in a second-hand uniform, was able to commandeer local troops and hold a local town hall to ransom. The trends were most extreme in France, the Dreyfus Affair demonstrating just how far militarism and anti-Semitism had penetrated politics. In Britain too, many believed in the necessity to inculcate martial qualities into society. Returning from observing the Russo-Japanese war, General Ian Hamilton remarked: “From the nursery and its toys to the Sunday school and its cadet company, every influence and affection, loyalty, tradition, and education should… impress upon young minds a feeling of reverence and admiration for the patriotic spirit.” It is often forgotten that Baden-Powell’s Boy Scout mantra of ‘be prepared’ was originally followed by ‘to die for your country’. Militarised societies were ripe for war.

One of the most compelling aspects of Europe’s descent into war in 1914 was the capacity of its political and military leaders to ignore inconvenient evidence about the changed nature of war. They thought
war would be short, fast and glorious, “no longer, certainly, than the war of 1870 that was consciously or unconsciously taken by that generation as a model.” But barbed wire was not patented until 1874, and the portable machine-gun ten years later. The romanticism of the age had idealised war as the acid test for individual bravery and national strength. Such delusions ignored the experiences of recent wars; the Russo-Turkish, Russo-Japanese and American Civil Wars had all exposed the destructive power of new industrial military technologies, particularly those designed for defence. Not everyone had been ignorant though. Socialists, like Jaurés, had predicted a catastrophe, and “Bloch’s ‘The Future of War’, foresaw a long drawn-out carnage.” Even as far back as 1890, the elder Helmuth von Moltke had recognised that Cabinet wars were a thing of the past, that in future, wars would require a total mobilisation of the nation. Epitomising the false optimism at the outbreak of war, the Kaiser told his departing troops: “You will be home before the leaves have fallen from the trees.”

One way to explain the ‘short war illusion’ is the predominance of the offensive in pre-war military strategic thought. Civilian and military elites were convinced that offensive action in war provided the decisive advantage. “All military authorities in Europe believed that attack was the only effective means of modern war, essential even for defence.” To help understand why, one must consider underlying contemporary ideas, such as Henri Bergson’s Élan vital (vital impetus). Evera argues that such virtues inspired a ‘Cult of the Offensive’, not just in pre-war France, but in all the European powers. It is this that fuelled the Darwinian armament race, both at sea and on land, and the tyranny of the intricate, inflexible war plans that carried the momentum in those final decisive days of July and August 1914.

Although Britain had effectively claimed victory in its naval race with Germany by 1911, the arms race on land accelerated as the two opposing alliance blocs strengthened. “By 1914 they were more evenly matched... both could now contemplate military action, this being probably the main respect in which the arms race made war more likely.” Parity of strength in 1914 should perhaps have deterred the belligerents from going to war, but the balance was ignored due to a faith in the offensive and, for the Austrians and Germans, a fear that time was running out. Both sides believed in a quick victory and shaped their respective plans accordingly. The Schlieffen-Moltke Plan, Russia’s Plan 19 and the French Plan XVII were all premised on early offensive action. This also helps explain why war could not be localised.

A. J. P. Taylor made the argument that military plans were directly responsible for the outbreak of war. He asserted, “the First World War [was]... imposed on the statesmen of Europe by railway timetable.” Mobilisation represented a serious escalation towards war, but the Schlieffen Plan combined the two and therefore reduced the diplomatic space for Germany. According to Taylor, the Russian decision to mobilise first exposed flaws in the German strategy, and forced the decision for war. “In 1914 [Schlieffen’s] dead hand automatically pulled the trigger.” Tuchman supported such a view: “Suddenly dismayed, governments struggled and twisted to fend it off... attempted to back away but the pull of military schedules pulled them forward.” Despite Tuchman’s apparent influence on President Kennedy’s decision-making during the Cuban Missile Crisis, the argument that military momentum held politicians hostage does not stand up. Unwieldy and inflexible military plans may have narrowed the options and choices but, as Stevenson argues, “the decision to resort to war was at bottom a political one.”

What is striking about the above analysis, in terms of the July Crisis, is the dysfunction in the civilian-military relationship. The lively eleventh hour debates in Russia and Germany, between political leaders and their chiefs of staff, demonstrated the acute lack of mutual understanding. Bethmann-Hollweg had no grasp of the mechanics of the Schlieffen Plan, and Moltke had failed to provide military options that could adapt to a range of political contingencies. In 1914, military plans were designed to execute specific and pre-conceived campaigns, not to support contingent political objectives. As Strachan argues, “the soldiers’ narrow political vision was matched by the remarkable military ignorance of the civilian leaders.” Militarism may not have caused the war directly, but its preponderance and privileged position in both the state and society helped shape the necessary conditions.
The Blame Game

“The European peace might have been a house of cards, but someone still had to topple it.”141

The analysis so far has illustrated that there are no quick or easy paths to deciphering the causes of war in 1914. An understanding of the outbreak of war requires not only careful consideration of innumerable and immediate contingent factors, but also a grasp of the powerful determining forces that had preceded it. Notwithstanding, the historical debate has long sought to attribute blame to one or more parties. This final section analyses the history and utility of the blame game. It seeks to highlight why the apportionment of responsibility has been important, and why the debate about it remains unresolved. In so doing, it will expose the historiographical evolution that has enabled a deeper, though not yet conclusive, understanding of the war’s origins.

Attribution of responsibility began before the war did. “Even before the first shots were fired, Europe’s statesmen were intent on weaving narratives designed to portray themselves as innocents responding to rival predatory aggressors.”142 As has been argued earlier, leaders could only harness the support of national populations by claiming victim status, but by 1919 the purpose of blame had evolved to the issue of guilt. Article 231 of the Versailles Treaty forced Germany and her allies to accept full responsibility for the war. To the victors went the spoils, reparation and political reprieve. This first round of the blame game was politically driven and it would not be the last. As Mombauer points out, “the argument over war guilt and responsibility for the war was at all times closely linked to the foreign and domestic policy concerns of the nations involved.”143

Germany’s vehement denial of culpability and desire to prove its innocence, led to a revisionist campaign during the interwar years. The trading of accusations between the belligerent powers led to an unprecedented, but selective publication of official documents and archives. In an attempt to discredit the former Tsarist regime the Bolsheviks were the first to make disclosures, with Germany, Britain, and France following. The volumes of document collections enabled a range of detailed accounts to be formed, both revisionist and anti-revisionist. These included the work of Wegerer, Barnes, Fay, Kantorowicz, Renouvin, Schmitt, and Albertini. A deluge of available source material resulted in a number of different accounts, distinguished by interpretation of the evidence.

With the passing of the interwar years, and the cumulative impact of the origins scholarship, “allocating blame became a less pressing concern… and the moment had come for a more conciliatory consensus regarding the war guilt question.”144 Whilst opinions over culpability remained divided, they converged towards a growing consensus that structural flaws in the international system were as much if not more to blame. In part, Woodrow Wilson’s drive to establish the League of Nations was an attempt to extinguish secret diplomacy; a practice he believed had helped to cause the war. Writing his war memoirs in the 1930s, Lloyd George famously concluded that, “the nations slithered over the brink into the boiling cauldron of war without a trace of apprehension or dismay.”145 The appeal of this inadvertence thesis is one that has endured, epitomised today by Christopher Clark’s Sleepwalkers.146

The comfortable consensus of the 1930s endured as the orthodox view for many years, the horrors of the Second World casting a shadow over both the July Crisis debate and the Treaty of Versailles. By 1955 it was believed the history of 1914 had been put to bed, that “in all areas the historian walks on safe ground.”147 It was a belief short lived. In 1961 the debate was reawakened and dramatically re-defined by Fritz Fischer, an extreme anti-revisionist. His books ‘Griff nach der Weltmacht’ (grasping for world power) and ‘Krieg der illusionen’ (The war of illusions) reasserted German war guilt. Drawing on extensive primary research, including archives from East Germany, Fischer argued, “Germany’s decision-makers had aggressive war aims in 1914 which they pursued throughout the war, and that Germany bore the main share of responsibility for the outbreak of the First World War which it unleashed in order to achieve these aims.”148 Fischer made the case that Germany’s strategy in 1914 was knowingly high-risk, that it had been pursued to ameliorate Germany’s social tensions and, by implication, that Nazi strategy in the late
1930s was simply the extension of an established policy. Strachan highlights how such arguments, "about Sonderweg and the ongoing preoccupation with the effect of Hitler – concerned continuity in German history, and therefore was as much about Germany’s current identity as about its past." Once again the history of the First World War had become entangled with contemporary political issues.

The highly controversial aspects of Fischer’s work stemmed primarily from his analysis of Bethmann-Hollweg’s September Programme of 1914, and the Kaiser’s War Council meeting in December 1912. He argued that civilian and military leaders in Germany had long colluded over a pre-emptive, annexationist war. “According to Fischer’s views, the policy that led to the outbreak of war was not one of blunders, but of design.” Critics, including Egmont Zechlin and Gerhard Ritter, attacked the retrospective basis of the September Programme argument, and the relative importance of the December War Council in 1912. Years of ferocious debate, and national soul searching followed Fischer’s claims, but his pioneering work profoundly influenced a new consensus that Germany had willed, or at least had risked war in July 1914, and was therefore primarily responsible for it. Many historians still reflect the less extreme aspects of Fischer’s work today. For example, David Stevenson agrees “that a preventative war was increasingly attractive to the German leaders from approximately 1911 but not decided on until July 1914.”

One of the more obvious weaknesses of Fischer’s thesis though was the lack of a comparative approach. His conclusions had reasserted German war guilt without any parallel review being conducted for the other belligerents. This issue was recognised by James Joll. His conviction on the issue helped shape his Stevenson inaugural, which appealed for a broader approach to historical research. In his introduction to the English language version of Griff nach der Weltmacht, Joll argued:

“Even if Fischer's work reinforces the belief that the German leaders bear the greatest weight of responsibility for the outbreak and prolongation of the First World War, it therefore imposes all the more strongly on British historians the duty of looking again at the record of the British Government.”

In recent decades, the First World War scholarship has increasing sought to address the gaps in the debate. The result has been a partial revival of the old orthodox view of the 1930s. A conference held in 2011, to mark 50 years since the initial publication of Fischer’s work, sought to review the current state of consensus within a comparative framework. It “revealed how fractured that consensus had… become.” In his summary of the conference, Steinberg wrote, “Much revision and revisionism marked these panels and a very different assessment of ‘war guilt’ began to emerge. The Germans look less guilty; the others conspicuously more.” Steinberg highlighted a number of observations. First, the debate had long been unbalanced, between the Fischer approach of seeking evidence that impugned Germany, compared to the British and French approaches, which tended to assume innocence. At the time Fischer was writing, the French had little interest in 1914 history and the Soviet leadership had no time for objective middle class historians. Austrian policy now appeared to be better understood, with new scholarship highlighting greater links between the Serbian government and terrorism. Likewise, the assassination of Stolypin in 1911 has been linked to a rise of Hawks in the Russian government, a development that made it more willing to risk war in 1914. In a reassessment of French judgments, it was noted that the immediate decision in Paris to support Russian actions in the Balkans, has received far less attention than the Schlieffen Plan in terms of its effect in widening the war. On reflection, Steinberg concluded, “the German ‘guilt’ in the causes of the First World War had been well and truly revised.”

Although historians now work from a much broader pool of historical source material than had been available in the past, there still remains little consensus in their revisionist conclusions. Building on the work of Norman Stone, McMeekin has tried to emulate Fischer’s approach, but in relation to Russia. Although he has been criticised for a tendency towards “tendentiousness and over-statement,” McMeekin usefully addresses the imbalance of belligerent scrutiny. In his own words, “the war of 1914 was Russia’s war even more than it was Germany’s.” McMeekin refocuses the debate on the war’s point of origin in the Balkans, and re-frames it in terms of being a war of Ottoman accession. The cause of Pan-Slavic solidarity is subordinated by high politics. McMeekin highlights a number of issues, such as the convenient loss of key correspondence between St Petersburg and Paris from July 1914, and the impact
on Russia of the shadow naval arms race that Turkey (with British support) had initiated in the Black Sea. He makes many more illuminating points, all reinforcing the argument that Berlin was not alone in having motives and a desire to risk war in 1914.

The same applies when the other belligerent motivations are scrutinised. Fischer may have largely ignored the role of Austria in his work, but few would now deny that the actions of Austria-Hungary initiated the crisis. As Herwig argues, “the initiative for war lay in Vienna. Habsburg and not Hohenzollern decided to settle accounts by military rather than diplomatic means. Both the direction and the pace of the July Crisis were dictated by Vienna… Vienna first resolved for war, sought German assurances, and then exploited them once received.” At the Fischer conference in 2011, Kronenbitter highlighted how the Fischer thesis had been considered ‘convenient’ for neutral Austria during the Cold War. Only Fritz Fellner stands out in seeking to ask more searching questions about Vienna’s role.

As in Austria, the Fischer polemic also failed to replicate itself in France. Keiger reminded the conference how stunted the French historiography has been, how it still remains trapped in the anti-German mould established by Renouvin. While France is the only European power that could legitimately claim to have fought a defensive war, the questionable Entente collaborations with Russia must surely be sufficient to warrant a renewed look at the responsibility of Paris?

Britain too has come under attack from historians. The first accusations came from the Kaiser before war began: “England alone bears the responsibility for war or peace, not we any longer!” This stark view stemmed from a German belief, at the time, that Britain had the choice to stay out of the war, and the power to influence French and Russian positions in the crisis, but chose not to. The basis of the argument is one taken up by Ferguson. He controversially makes the argument that Britain could and should have stayed out of the war, but in specific relation to the causes of the war itself, he argues, “the uncertainty about Britain’s position probably made a continental war more rather than less likely, by encouraging the Germans to consider a pre-emptive strike.” It has been a divisive argument among British historians, not least given the consequences of the war itself.

It is certainly possible that a clearer indication of Britain’s position earlier in the crisis may have helped to deter war, but this argument fails to allow for the lack of unity in the British Cabinet until the German attack on Belgium. It also underplays Grey’s diplomatic predicament, one that Ferguson himself acknowledges: “Grey’s strategy in trying to turn the ententes with France and Russia into quasi-alliances had been to deter Germany from risking war. However, now he feared that too strong a signal of support for France and Russia… might encourage the Russians to do just that. He found himself in a cleft stick: how to deter Austria and Germany without encouraging France and Russia.” Grey was responsible for Britain’s confused position, straddling between isolation and alliance, a situation further complicated by a number of military commitments Grey agreed with France, notably the 1912 Anglo-French Naval Convention. Additionally, however much London may have condemned Germany for invading Belgium, it is interesting how little it had to say about Russia’s move on Vienna. Standing up for the liberty of smaller states may have been important to the British in 1914, but the Belgian Casus Belli masked grand strategic self-interest. Such logic cannot place Britain’s hands on the 1914 trigger, but it is harder to distinguish Britain’s blundering and high-risk foreign policy from that being pursued by Berlin.

Nations blundering into war in more general terms is precisely where Clark concludes his latest book. For Clark, the ‘blame game’ serves no historical purpose and should be avoided. Indeed, it dangerously risks casting international relations into binary right and wrong camps. To illustrate his point Clarke questions, “were the Serbs wrong to seek to unify Serbdom? Were the Austrians wrong to insist on the independence of Albania? Was one of these enterprises more wrong than the other?” Clarke makes a strong case against ‘prosecutorial narratives’, arguing: “The quest for blame predisposes the investigator to construe the actions of decision-makers as planned and driven by a coherent intention.” Given the raft of myths, controversies, and contradictions in the 1914 origins literature, Clark’s revisionist argument is compelling. Yet still it remains unsatisfactory. Inevitably, some historians have attacked Clark’s account as ahistorical, that it ignores the evidence. Whilst Strachan praises Clarke for “getting closer to how most statesmen saw war’s place in international relations in 1914 itself,” he takes issue with the
depiction of Europe’s leaders as ‘Sleepwalkers’. Strachan contends, “to call the principal actors sleepwalkers is to suggest that they were unaware of what they were doing, had little conception of war or its dangers, and were not seized of the seriousness of what they were about. The bitter irony is that they knew enough about all these things but, although awake to them, could still not control them.” As the centenary approaches, it seems the riddle of the origins debate is far from solved and is likely to go on for some time yet.

**Conclusion**

Having considered the various interpretations of the July Crisis, the impact and influence of the broader underlying context, and the interpretations and judgement of history, it is the outlook and behaviour of Europe’s leaders, the ruling elites of 1914 that stands out most. Beyond all the complexity, contingency and contradiction, it is the character of key individuals in each capital that stymies rational action throughout. In concluding her recent book, which follows a similar logic to Clark’s, MacMillan also shies away from attributing blame, but not without hinting at where it may lie. She suggests, “perhaps all we can hope for is to understand as best we can those individuals, who had to make the choices between war and peace, and their strengths and weaknesses, their loves, hatreds and biases.”

It is accepted that all the leaders suffered from miscommunication and made miscalculations during the crisis, Germany more than most. But it was more than this. What stands out is the influence of character, of attitude, of behaviour. Joll’s unspoken and, in some cases, spoken assumptions return to the fore. In most accounts that depict the decision-makers of 1914 one can invariably detect arrogance, incompetence, paranoia, a lack of empathy, weakness, and above all, egotism. As Stoessinger argues, “there was a pervasive tendency to place the preservation of one’s ego before the preservation of peace. There was little insight and no vision whatsoever…. It was not fate or Providence that made these people fail so miserably; it was their own evasion of responsibility.” The ruling classes had all been schooled in aristocratic codes of behaviour, and inculcated with class values. In 1914, “the issue was not that nobody wanted war, but that everybody was willing to risk it – and if necessary to fight rather than give way.” Misplaced codes of honour and deep prejudices compromised the sophisticated leadership that July 1914 demanded. Missing too was the intellect necessary to navigate away from the abyss. As Joll articulates, “the leading European statesmen… were swept away towards a goal they could not see along a road for which they lacked the maps.” It was this that sealed the fate of millions of workers across Europe. They were sent to defend the existing order, ready victims of privileged elites, who “sought to solidify their position in the great struggle against the forces of change.” This was the decisive element that led to outbreak of war in 1914.

Like many assessments on the outbreak of war in July 1914, this one has chosen to apportion blame. It has done so consciously, recognising Joll’s point that, choice of cause “is conditioned by our own politics and psychological interests.” This paper contends those who stress a greater burden of responsibility for Berlin, Vienna, St Petersburg, or elsewhere, identify their smoking gun in too narrow terms. Germany and Austria’s actions in July 1914 may have forced the crisis, but the issues at the root of the problem were international. In responding to the crisis, the Entente powers simply moved to shore up their own interests at the expense of the Central powers. This cannot justify the one-sided guilt clause that followed the cataclysm, and the century of debate that followed.

The outbreak of the First World War was a failure in the application of reason, and the failure rested with the governing classes. “If there is one lesson that stands out… it is the almost unbelievable blindness, tenacity, cruelty, and unscrupulosity with which the governing classes cling to their privileges and power at any cost to their suffering peoples and to the wider interests of peace and civilisation.”

**Endnotes**

3 Ibid.
3 Ibid.
8 Clarke, *The Sleepwalkers*, xxvi.
9 MacMillan, *1914 and 2014: should we be worried?*, 60.
10 Ibid.
11 MacMillan, *1914 and 2014: should we be worried?*, 60.
16 Ibid.
19 Dimitrijević had been responsible for the assassination of the Serbian King Alexander and his wife Queen Draga.
21 Clarke, *The Sleepwalkers*, 41.
29 Ibid., 91.
30 Berghahn, *Germany and the Approach of War*, 201.
31 Ibid., 198.
32 Ibid., 199.
34 Berghahn, *Germany and the Approach of War*, 205.
35 Ibid., 208.
39 Berghahn, *Germany and the Approach of War*, 212.
40 Ibid., 209.
41 Ibid., 212.
42 MacMillan, *The war that ended peace*, 539.
44 Ibid., 59.
48 Ibid., 73.
52 Mombauer, *The Origins of the First World War*, 16.
55 Ibid., 41.
60 Joll, 1914: the Unspoken Assumptions, 328.
69 Not including the German wars of unification, which did not involve all the Great Powers.
72 Clark, Globalization and Fragmentation, 41.
73 Ibid., 40.
77 Kissinger, Diplomacy, 201.
79 Strachan, The Outbreak of the First World War, 125.
80 MacMillan, The war that ended peace, 50.
82 Ibid., 199.
83 Ibid., 244.
84 Ibid..
86 Joll, 1914: the Unspoken Assumptions, 309
87 Stevenson, Learning from the past: the relevance of international history, 17.
88 Strachan, The Outbreak of the First World War, 116.
89 Clark, Globalization and Fragmentation, 42.
91 Ferguson, The Pity of War, 22.
92 MacMillan, The war that ended peace, 11.
93 Tuchman, The Proud Tower, xiv.
94 Strachan, The Outbreak of the First World War, 141.
95 Ibid., 141.
97 Gordon, Domestic Conflict and the Origins of the First World War, 94.
98 Strachan, The Outbreak of the First World War, 169.
99 Joll, 1914: the Unspoken Assumptions, 318
100 Clark, Globalization and Fragmentation, 51.
102 Ibid., 98.
103 Tuchman, The Proud Tower, xiv.
104 Ibid.
105 Joll, 1914: the Unspoken Assumptions, 321
106 Ibid., 322.
107 Strachan, The Outbreak of the First World War, 173.
109 Ibid., 238.
110 Strachan, The Outbreak of the First World War, 179.
111 Ibid., 177.
112 Joll, 1914: the Unspoken Assumptions, 324.
113 Joll and Martel, The Origins of the First World War, 279.
114 Joll, 1914: the Unspoken Assumptions, 324.
115 Ibid., 326.
116 Tuchman, The Proud Tower, 462.
117 Strachan, The Outbreak of the First World War, 175.
119 MacMillan, The war that ended peace, 249.
120 Strachan, The Outbreak of the First World War, 232.
122 Strachan, The Outbreak of the First World War, 237.
124 Neiberg, Dance of the Furies, 5.
125 Strachan, The Origins of the First World War, 431.
126 Clark, The Sleepwalkers, 214.
128 MacMillan, The war that ended peace, 259.
130 Stevenson, The Outbreak of the First World War, 56.
131 Stephen Van Evera, Causes of War (New York: Cornell University Press, 1999), 204.
132 Term used by Lancelot Farrar, The Short-war Illusion: German Policy, Strategy and Domestic Affairs, August-December, 1914 (Oxford: ABC-Clio, 1973)
134 Van Evera, Causes of War, 194.
135 Stevenson, The Outbreak of the First World War, 50.
137 Ibid., 20.
140 Strachan, The Outbreak of the First World War, 123.
143 Mombauer, The Origins of the First World War, 54.
144 Ibid., 105.
145 Ibid., 95.
146 Clark, The Sleepwalkers.
148 Ibid.
149 Strachan, The Outbreak of the First World War, 6.
150 Mombauer, The Origins of the First World War, 128.
151 Ibid., 133-138.
152 Stevenson, Learning from the past, 15.
153 Mombauer, The Origins of the First World War, 192.
156 Ibid.
158 Ibid., 248.
160 Strachan, The origins of the First World War, 436.
162 Mombauer, The Origins of the First World War, 190.
166 Ferguson, The Pity of War, 80.
167 Ibid., 155.
168 Clark, The Sleepwalkers.
169 Ibid., 560.
170 Ibid.
173 Ibid., 435.
174 MacMillan, The war that ended peace.
175 Ibid., 605.
176 Stoessinger, Why Nations Go to War, 21.
177 Joll, 1914: The Unspoken Assumptions, 315.
178 Stevenson, Learning from the past, 16.
180 Steinberg, “Old Knowledge and New research: A Summary of the Conference on the Fischer Controversy 50 years on,” 249.
‘Assess the impact on the Allied war effort of Germany’s V-weapons offensive. To what extent is this impact reflected in the current implications of Weapons of Mass Effect?’

Group Captain M R JOHNSON

"Who can think without horror of what another widespread war would mean, waged as it would be with all the new weapons of mass destruction"¹

The Times, 28 December 1937

Introduction

Between June 1944 and March 1945, Germany launched 15,500 V-1 ‘flying bombs’ and V-2 rockets at targets in England and Europe.² While the V-weapons offensive did not succeed in meeting Hitler’s aspiration of altering the course of the war by attacking the morale of the population,³ this belies their broader effectiveness as weapons that delivered significant strategic influence. The Germans’ overall strategy for the use of the weapons was incoherent, and included disagreement as to whether they should be used against the civilian population or as weapons to strike at military targets.⁴ Nonetheless, the indiscriminate nature of the weapons resulted in an impact that outweighed their capabilities as military weapons. This effect was magnified by the threat that they might be used to deliver chemical and biological agents.

The broader impact and influence of the V-weapons has endured beyond WWII. As the forerunners to cruise and ballistic missiles, the V-1 and V-2 marked the introduction of the use of missiles and rockets to deliver strategic influence and, as such, have provided a template for the contemporary implications of Weapons of Mass Effect (WME). British Doctrine defines WME as “weapons capable of a high order of effect…beyond the traditional lethal domain”.⁵ By definition, therefore, WME include all previous weapons that could be categorised as Weapons of Mass Destruction (WMD), such as nuclear, chemical, biological and radiological. A key area of difference, however, is the emphasis on the ability of WME to deliver psychological effect and, therefore, their ability to influence the cognitive domain as well as the lethal. In assessing the implications of modern WME, the paper will consider cruise and ballistic missiles with the potential to deliver traditional WMD warheads. As observed by former US Secretary of Defense William Perry, such weapons in the hands of rogue or 3rd world states could “constitute the greatest single danger to…world security.”⁶ While recognising their overall importance in the debate, the paper does not specifically consider implications of the threat posed by nuclear weapons, or the unique nature of the evolving implications posed by cyber warfare as a potential WME.

By examining historical analysis of the V-weapon offensive, the paper shows that the psychological impact of the threat posed by the weapons had a major influence on strategic thinking, and had a tangible impact through the diversion of resources to counter the threat. It will conclude, however, that this did not have a decisive effect on the outcome of the war. Extrapolation from more recent campaigns highlights many parallels between the impact caused by the V-weapons and the implications of modern WME, particularly in terms of strategic considerations. Finally, the paper will argue that the implications associated with allocating assets to counter today’s threat is likely to be more complex and challenging than at any previous time due to pressures and imperatives caused by resource and fiscal constraints.
Psychological Impact

Although Hitler did not succeed in his objective of destroying the will and undermining the support of the English population, the fear caused by his ‘Vengeance’ weapons undoubtedly affected morale and had a psychological effect that impacted allied considerations for the remainder of the war. The first V-1 landed in London on 13 June 1944; this was followed by the first V-2 on 8 September 1944. Between June 1944 and April 1945, more than 8,600 V-1 and V-2 weapons attacked England; nearly 1.5 million people had evacuated London by September 1944 and, by the end of the war, 24,165 had been killed or seriously injured by the V-weapons. Although the V-1 and V-2 both brought terror and fear, they did so in different ways. The V-1 was an inherently inaccurate weapon which was used indiscriminately by the Germans. The fear this engendered was enhanced by the fact that its distinctive droning noise became silent once the engine cut, giving a terrifying notice of the destruction that was to follow. Although more accurate, the V-2 was silent and therefore gave no notice prior to impact. While this may have engendered a “fatalistic attitude”, the V-2 also marked a shift in the way civilians could be targeted; Coblenz describes this effect as “the introduction of chaos… into human affairs.”

There was strong concern that the psychological impact of prolonged exposure to the V-weapon attacks would lead to a loss of public support, with calls for the British Government to seek a peaceful settlement of the war. A further consideration concerned the impact that the V-weapon offensive had on the morale of those soldiers engaged in Normandy, who were distracted and worried for the welfare of their loved ones at home in England. Deliberate efforts were taken to mitigate this risk, including through the use of the press. This was reflected in an article in The Times in June 1944, which suggested that “The aim of these nuisance raids is no doubt to shake the morale of the British public, which has never been stronger than today.” This was one example of the leadership and authorities playing down the level of fear, and the psychological effect of the V-weapons, as a way of maintaining the support of the people.

But there is plenty of evidence to suggest that the V-weapons did have a significant effect on the morale of the population. Johnson highlights the level of fear the V-weapons caused by quoting a London resident; “The flying bombs were the terror of our lives…we sat under the table with our hearts in our mouths until the dreadful explosion came.” Furthermore, the overall psychological effect and impact on morale was formally recorded by Squadron Leader Herbert Bates, in a 1945 (although not released until 1994) study for the Air Ministry which stated; “The toll of death, injury and damage to property from the flying bomb attacks was greater than anyone imagined”. He went on to add; “In reality it did a great deal to morale.” Although not used, it was also believed that the V-weapons had the potential to deliver chemical and biological agents. It is likely, therefore, that passive defence measures, such as the issuing of gas masks, would have increased the psychological effect and level of fear of the weapons. This threat also affected the political and military leadership, who were immensely relieved to discover that the V-weapons used against England did not carry chemical and biological weapons. The attacks also caused a mass exodus of workers from London; those that remained were terrified and exhausted, and stayed away from work to deal with damage to their own properties. As a consequence, it is estimated that the V-weapon offensive resulted in war production in London being reduced by 25%. Despite this reduction in capacity, however, it is unlikely that the attacks would have materially affected the war effort because of the proximity of the end of the war in 1944-45. Moreover, the attacks would not have affected the industrial effort of the other Allied powers.

More recent events have shown that WME are likely to be used by an adversary as a weapon of strategic influence, by capitalising on the fear and psychological impact that the asymmetric use of such weapons would cause. Pastel and Ritchie describe these weapons as “effective agents of terror”, and this would be borne out by the devastating effect of Iraq’s use of chemical weapons...
in its war with Iran in 1980-1988 which caused over 30,000 casualties. The psychological effect of Saddam's use of WME during this war was significant. As well as witnessing the horror, the impact on morale caused by publishing the medical effects of the chemical weapons was such that volunteers for Iran's Revolutionary Guard fell by one third. Moreover, the fear that Saddam's Scuds would be used to deliver chemical weapons, during the 'War of the Cities', also reportedly resulted in up to one half of Tehran's population evacuating the city.

Domenici argues, however, that chemical and biological weapons do not have to be used to cause fear and have a psychological effect, rather, their potential use could be sufficient. This was very much the case with Saddam's use of Scud missiles in 1991, which was intended to alter the course of the war through their psychological impact. Although he did not use chemical and biological weapons on this occasion, the threat that they might be used had a strong psychological effect on Israelis and resulted in gas masks being issued to the civilian population.

Therefore, through delivering psychological impact, WME have proven to be effective weapons that could be used by a potential adversary to achieve asymmetric advantage by attacking civilian targets. The extent to which the V-weapons, and the threat that they could carry chemical and biological agents, affected the strategic considerations of the Allied leadership will now be explored, alongside related contemporary strategic implications.

**Impact on strategic considerations**

In order to minimise the risk of the V-weapons being effective in targeting civilians, the highest political priority was placed on pre-empting and neutralising the V-weapon threat before the first weapon could be launched. As King and Kutta argue, it would have been unacceptable for the civilian population to suffer again as they had during the Blitz. The V-weapon threat was exacerbated by reports of Hitler's chemical and biological programme. This became an imperative in strategic decision making as the Allies embarked on a massive programme for developing their own chemical and biological weapons, because they had to "prepare for the worst". Although intended for retaliation, Churchill's serious consideration of the pre-emptive use of these weapons, to counter Hitler's threat, highlights the influence that the V-weapons had on strategic decision making.

The imperative to counter the V-weapon threat also adversely influenced strategic planning decisions. For some time, Montgomery had been unsuccessful in securing Eisenhower's support for Operation Market Garden. As the V-2 threat became clearer, Montgomery re-submitted the plan to include the opportunity to neutralise the rocket threat. Not only did Eisenhower agree to the plan, but he allocated it the highest priority. This was ultimately a flawed decision as the operation was a failure. D'Este describes Eisenhower's decision to agree to the Market Garden plan as one made "more from a sense of...pressure to overrun and put out of commission Hitler's V-weapon sites in Holland than from a solid military foundation." Therefore, although the primary objective of Market Garden was not the V-weapons, it became the decisive imperative. The urgency of the requirement to neutralise the V-weapon threat therefore clouded military judgement, and was the influencing factor that led to Eisenhower agreeing to the ill-fated plan.

The threat of the V-weapons also brought to the fore national imperatives and considerations within the Allies, with the inevitable potential for friction. In evaluating the most appropriate means to counter the V-weapon threat, the US proposed that a joint US/British committee should replace the British Air Ministry in having responsibility for countering the threat. This was refused by the British leadership in unambiguous terms, on the grounds that it was the British people that were under threat and, therefore, they would lead with the response. The US was also seen to follow national interests with regard to the threat. For example, a key driver behind US support for
the UK allocating a priority to the V-2 was that they believed that the rockets had the range to target mainland US. Intelligence sharing also became a source of friction. A report raised by a war committee, established in Washington to evaluate the implications of the rocket threat, highlighted discontent over a lack of intelligence sharing; it stated that the committee was “strongly impressed by the hesitancy of British leaders to reveal the true nature of the danger.”

Although these frictions were an issue, there is little evidence to suggest that they had a major impact on either the cohesion of the strategic alliance, or the overall war effort.

In any future conflict, our strategic centre of gravity is likely to be the cohesion of a coalition, with support of the people a critical requirement. Therefore, the need to act to minimise the psychological impact will be just as much of an imperative in contemporary considerations as it was during the V-weapon campaign, as shown by the coalition imperative to keep Israel out of the Iraq war in 1991. Recent events in Syria have also highlighted that WME will continue to be exploited by potential adversaries, to introduce frictions in creating and maintaining cohesion within a coalition. In 2012 a Syrian Foreign Ministry spokesman stated that, although his government had no intention of using chemical weapons, it might consider doing so if “Syria faces external aggression”. The intent behind this statement was to use WME to introduce uncertainty into the minds of potential coalition partners and, thus, make creating coalitions more difficult; a point reinforced by Dominici in his discussion regarding the coalition in Iraq in 2003. The discovery of a WME threat during an operation could also present strategic challenges. National imperatives would then determine how a coalition partner responded, potentially leading to a withdrawal of support or military contribution. The implications that WME presents for a coalition could therefore be significant, with particular emphasis on the imperative to ensure that its cohesion is not undermined.

Ranger and Wiencek argue that the presence of WME will be a key factor that complicates strategic decision making, and this will include challenges associated with deterrence and potential responses. While the Allies’ chemical and biological programme may have deterred Hitler from use of his programme, a combination of ethical considerations, and adherence to jus in bello principles and treaty obligations quite rightly precludes this from being an appropriate form of modern deterrence for the West. Credible messaging, therefore, could be extremely important. President Obama’s statement in 2012, aimed at deterring Syria from using WME, caused much debate and could potentially have wider strategic implications. In this instance the US did not follow through with a specific response despite the warning that, “…a red line for us is we start seeing… chemical weapons moving around or being utilized... That would change my equation.”

The decision to go to war with Iraq in 2003 also highlights the challenges that WME presents concerning intervention. The decision was taken because of the perceived need to counter the threat of Iraq’s WMD, however the belief that Saddam had this capability at his disposal was based on ambiguous intelligence linking a chemical and biological threat with ballistic missiles. As with the experience of Market Garden, history shows that the imperative to counter this perceived threat clouded political and strategic level judgement and decision making. These examples highlight that the potential presence of WME could continue to have a significant, and sometimes detrimental, impact on strategic considerations.

Although the threat of the V-weapons did have an effect on strategic thinking for the Allied leaders in WWII, there is no evidence to suggest that this had a major impact on the overall Allied effort. Experience from more recent case studies would indicate that the imperative to counter the threat of WME could continue to lead to strategic challenges. The paper will now explore the resource implications associated with countering the V-weapon threat, and its contemporary parallels.
Impact on resources

There is a strong argument that the greatest impact the V-1 and V-2 weapons had on the overall war effort was in the level of resources that were diverted to counter the threat. As Collier identified, “The great question was whether a substantial part of the Allied bomber effort should be switched from the battle in Normandy or the bombing of Germany to the rather daunting task of knocking out … sites”. The majority of these resources were allocated to Operation Crossbow, the Allied effort to coordinate pre-emptive aspects of dealing with the V-weapon threat. This included the intelligence gathering contribution and offensive strikes against the range of targets associated with the V-weapons. There was general agreement amongst the Allied leaders that air power would be the key to countering the launch of the V-1 and V-2 weapons; this was predicated on the fact that the Germans still had control of continental Europe in 1943. But this could only be achieved if sorties were diverted from other missions. From 17 August 1943, when the RAF conducted the initial raid against the Peenemunde rocket development complex, in excess of 6,000 bomber missions were diverted to pre-emptively attack Crossbow targets before the V-weapon offensive started in June 1944. In January 1944 alone, 38% of all missions were assigned to meet Crossbow requirements. This would reinforce Joseph Angell’s observation of the challenge as “a diversionary problem of the first magnitude.”

With such a high percentage of missions diverted to counter the V-weapon threat, it was inevitable that such action would lead to frictions. Specifically, during the period 1943–1945, Operation Crossbow was competing with Operations Pointblank and Overlord for the same resources. Particular issues of concern were raised by both Air Chief Marshal Harris and Lieutenant General Spaatz, who didn’t want resources diverted from Pointblank, the offensive bombing campaign over Germany, to the defensive Crossbow campaign. Central to this disagreement was a failure of the operational commanders to grasp the political imperative to neutralise the V-weapon threat and minimise any possible risk of losing the support of the population. Eisenhower, however, did understand the political imperative and, with overall command for all bombing assets, he prioritised Crossbow missions over all other missions, except those that met, in his words, “the urgent requirements of the battle”. This was supported by Churchill. As the diversion of sorties to Crossbow became an enduring friction, Churchill personally intervened in July 1944 when he declared that “Subject to the overriding needs of the Battle of France, all Britain’s available resources must be used to try to counter the flying bombs.”

While the two V-weapons presented very different challenges, in terms of counter-measures and the associated resources that needed to be allocated, their collective “randomness” added to the resource burden necessary to counter the threat. The nature of the V-2 rocket was such that it could not be defended against once airborne, therefore, destroying the launch sites became a critical requirement. The V-1 on the other hand could be intercepted with responsibility delegated to the Air Defence of Great Britain. Again, significant resource was required for the associated three lines of defence: multiple squadrons of fighters including Tempests, Spitfires and Meteors provided the first line of defence; the second line was provided by anti-aircraft guns; and the third line by barrage balloons.

Despite Harris’ claim, in July 1944, that diverting resources to Crossbow had contributed to undermining most of Bomber Command’s efforts over the preceding 3 years, there is insufficient evidence to suggest that it delayed the outcome of the war, particularly considering the numbers of aircraft available to the Allies. The conclusion reached by the United States Strategic Bombing Survey is that the diversion of resources had an insignificant impact overall on the Allied war effort. Kipphut reinforces this, and points to the fact that the Crossbow missions successfully delayed the use of the V-weapons by 3-6 months, which was long enough to enable the Normandy landings to take place as scheduled. There is also an argument that Hitler’s
persistence with the V-weapons, and his obsession with the V-2 in particular, had a positive impact on the Allied war effort. In order to pursue the V-2, Hitler ordered that all available resources should be allocated to the programme. This resulted in the cancellation of projects, such as the Wasserfall anti-aircraft programme, with the resultant effect that there was no need for dedicated suppression of enemy air defence missions, because allied bombers encountered significantly less German defences during their bombing raids. While the V-weapon campaign was successful in meeting Hitler’s objective of diverting resources, this did not have a decisive effect and did not have the desired impact of altering the course of the war. In particular, it came too late to delay the Normandy offensive.

It is likely that one of the most significant contemporary implications raised by WME would also be associated with the allocation of resources to counter the threat. The significant level of resource diverted to counter the threat of Scuds being launched at Israel in 1991 (1,500 strike sorties) mirrored that allocated to counter the V-weapon threat. But Kipphut makes the point that future challenges will be greater because the threat is likely to comprise more technologically advanced ballistic and cruise missiles. A mass of resource alone will no longer be sufficient, or indeed be available, to counter this threat, which will result in the need for a more sophisticated approach and a prioritisation of assets.

The threat that these weapons could carry chemical, biological and nuclear warheads is an additional imperative in the requirement for comprehensive defences. But countering the threat of these WME needs to go much deeper and, according to a senior Pentagon official, should comprise a layered approach, including prevention, and active defence, both of which will have implications for resource allocation. While prevention will require international community enforcement of regulations and treaties, rogue states may not necessarily pay much heed to this. Iraq openly flaunted its disregard for the international community’s efforts during the Iran-Iraq war, with Syria threatening to do the same in 2013. Another facet of prevention, therefore, is the possibility of offensive pre-emptive strikes. But pre-emptive strikes would not be acceptable in all instances and, as discussed above, would the appropriate resources exist for a successful campaign?

Amongst the range of counter-measures available, it is possible that defensive measures could be the most complex. Ballistic missile defence systems are at the forefront of these considerations. Given the challenges associated with developing a comprehensive defensive system, it is no surprise that the US is fostering a network of close partnerships, including with NATO, Japan and Israel. But at a time of restraint in defence spending, prioritisation in the allocation of scarce resource will be essential and will not be without friction. For example, the US announced in March 2013 that it would divert its missile defence efforts from Europe to the US west coast to counter the developing North Korean ballistic missile threat, but this could only be achieved by cancelling its proposals to deploy a similar capability in Europe. Therefore, while the contemporary threat posed by WME has also been shown to require the diversion of resources, the future challenges are likely to be more complex, with the potential for greater friction internally and amongst partners in determining the most effective means of allocating appropriate resources to counter the threat.

Conclusion

The paper has shown that the V-weapons did not, as Hitler had hoped, impact on the Allied war effort sufficiently to alter the course of the war. They did, however, have a significant effect which extended beyond military considerations. As terror weapons, and underpinned by the belief that they could be used to deliver chemical and biological agents, the psychological impact they had on both the population and war leaders had a major influence on strategic considerations.
throughout the later stages of the war, particularly with regard to the diversion of resources. Paradoxically, and rather than limiting their effectiveness, the inaccuracy and indiscriminate nature of the V-weapons reinforced their potential categorisation as original weapons of mass effect.

Analysis has also shown that there are many parallels with contemporary implications of WME. In particular, evidence suggests that the psychological impact of these weapons could remain significant in influencing strategic considerations, especially those associated with protecting the centre of gravity. Not all historians, however, agree with the paper’s thesis regarding the legacy of the V-weapons as WME. For example, Neufeld argues that “The ballistic missile wasn’t an effective weapon until you put a nuclear warhead on top of it – and suddenly it became a super weapon”. But this misses the point regarding cognitive influence. Masters supports this view by arguing that while ballistic missiles may still not necessarily be very accurate, their greatest impact will be in the psychological domain by targeting populous regions.

Finally, the statement quoted from The Times that introduces this paper is as valid today as it was in 1937. But even if a contemporary war is not ‘waged’ with WME, the threat posed by missiles with the capability to carry such warheads, even if not used, will ensure strategic influence by having psychological impact. The imperative to mitigate this impact will therefore remain strong and the enduring challenge will be associated with decisions regarding how to counter a possible WME threat. This is reflected in the fact that many senior interlocutors have recently highlighted the need for the UK and NATO to give serious consideration to robust ballistic missile defence measures. As the UK begins to refine its thinking on SDSR15 considerations, investment in a missile defence capability may be a key outcome that mirrors the priority placed on cyber defence considerations in SDSR10. While it is impossible to predict whether this will be a key outcome, and at what cost, it can be predicted, with a degree of certainty, that prioritising such a decision will not be without its difficulties.

Endnotes

7 United States Strategic Bombing Survey, V-Weapons (Crossbow) Campaign, (Military Analysis Division, 1947): 1
9 E Jones, R Woolven, B Durodie, S Wessely, “Civilian Morale During the Second World War: Responses to Air Raids Re-examined” Social History of Medicine, Vol. 17, no. 3 (2004): 473.
10 Stanton A. Coblentz, From Arrow to Atom Bomb: The Psychological History of War (London: Peter Owen Limited, nd), 434.
11 King and Kutta, Impact: The History of Germany’s V-Weapons in World War II, 211
15 Ibid., 64.
18 Johnson, V1 V2, 88.
26 Goldstein, "Pinpricks", 91.
27 Johnson, V1 V2, 138.
29 Goldstein, “Pinpricks”, 94.
31 Goldstein, “Pinpricks”, 81.
33 Domenici, “Countering weapons of mass destruction,” 146.
38 Collier, *V-Weapons*, 77
40 McFarland, “Preparing for What Never Came,” 111
41 Goldstein, “Pinpricks”, 87.
42 Ibid.
44 Goldstein, “Pinpricks”, 89.
46 Ibid., 190.
48 United States Strategic Bombing Survey, *V-Weapons (Crossbow) Campaign*, 36
50 Johnson, V1 V2, 164.
52 Kipphut, "Theatre Missile Defense," 47.
54 Ranger and Wiencek, *The Devil’s Brews II*, 44.
56 Ibid.
59 Including during HCSC presentations at PJHQ and NATO HQ.
A critical examination of the role of leadership in the handling of National Security crises: Charismatic leadership and self-reliant team paradigms.

Sq Ldr Sue Gray

The traditional view held by researchers (Weber 1947, House 1976) assumes that in a crisis “the successful leader must become decisive, demonstrate a ruthless ability to focus on the problem and to ignore the siren calls of the sceptics and the cynics” (Grint 2005 p.1468). However, academic discourse has since widened considerably and now encompasses a variety of different leadership and management roles to tackle crisis situations. The UK National Security Strategy document (Oct 2010) highlights the complexity surrounding National Security Crises, with threats potentially emanating from a “myriad of sources” and the “whole government approach” encompassing all instruments of national power together with elements of the private sector and the public (ibid p.5). The term ‘crisis’, defined as “a change, either sudden or evolving, that results in an urgent problem that must be addressed immediately...” (Harvard Business Essentials 2004 p.xvi), can be categorised as Critical, Tame or Wicked (Grint 2005, 2008); these can help determine and, indeed, legitimise, the leadership response evoked. Grint’s (ibid) typology will be used to examine whether National Security Crises are best handled by charismatic leaders, self-reliant teams, well trained officers, executives or lower grade individuals. The first three will be critically examined before an alternative view of leader as ‘sense-maker’ (Weick & Sutcliffe 2007) or ‘bricoleur’ (Grint 2008) is proposed to handle crises categorised as Wicked. Subsequently, executives and lower grade individuals will be examined collectively as contributory actors. It will be seen that, whilst crises have traditionally been seen to demand charismatic leadership, this is seldom the most appropriate approach. In addition, whilst each leadership response has some utility in handling different types of crisis, depending on how it is categorised, each has limitations which could potentially affect the effectiveness of the outcome. It will also be seen that the more complex the National Security Crisis faced and the more actors involved, the greater the need for wider engagement with experts at all levels, including executives and lower grade individuals.

The subjectivity of the definition of ‘crisis’ is highlighted by Grint (ibid p. 14), who proposes that a situation of ill-defined threat only becomes a crisis when that threat is pronounced as such by someone in authority and accepted as such by followers. As what constitutes a crisis seems to vary considerably depending on how the situation is framed, this impacts on the approach chosen to resolve it (Grint 2005). Grint (ibid) also differentiates between Critical, Tame and Wicked problems which he correlates with Command, Management and Leadership. A Critical problem is self-evident in nature, encapsulates very little time for decision-making and is, therefore, associated with the need for a Commander to take decisive action. A Tame problem, whilst it may be complicated, is likely to have occurred before and is resolvable through unilinear acts; as there is only a limited degree of uncertainty, Managers may be able to apply appropriate processes. A Wicked problem, on the other hand, is complex, often intractable, has no unilinear solution, no ‘stopping’ point and a high degree of uncertainty (ibid). As any attempt to solve it generates other problems and there are no clear causal links, a Leader is required to ask the right questions as opposed to providing answers. As situations can shift across the boundaries of Critical, Tame or Wicked (ibid), a National Security Crisis could fall into any of the three categories at different times. Similarly, although decision-makers will have a preferred response and may seek to frame situations in order to legitimise their preference, there is no one ‘appropriate’ form of response for a particular situation; it can switch between Command, Management and Leadership depending on how the situation is constituted (ibid). As Grint’s (ibid) basic typology provides a useful framework, it will be used to examine the relative merits of charismatic leadership, self-reliant teams and well-trained officers in handling National Security Crises.
Researchers, such as Weber (1947) and House (1976), have postulated that charismatic leadership emerges in times of crisis, as the stress involved causes followers to look to leaders to deliver them from their difficulties (Northouse 2010). In such situations, a leader with exceptional personal qualities is believed to arise, attracting followers who believe in the vision and who profess the leader to be extraordinary (Defence Leadership Centre 2004 p. A-15). Initially defined by Weber (op cit) as a special personality characteristic that gives a person exceptional powers, is reserved for a few, and is of divine origin, the subjectivity and important role of followers in validating charisma in their leaders was later appreciated (Bryman 1992, House 1976). Furthermore, being dominant, self-confident, having a strong desire to influence others, and a strong sense of one’s own moral values (House ibid) could be seen to be desirable in handling a crisis, especially during the early stages (Lewis 2006). Similarly, the ability to motivate, inspire and empower followers to succeed in times of uncertainty (ibid) could be seen as advantageous.

Less desirable effects of charismatic leadership, such as follower obedience, unquestioning trust in the leader’s ideology and alignment between followers’ and leader’s beliefs (ibid), and the tendency for charismatic leaders to be unnecessary risk takers who find cooperation with others “awkwardly bureaucratic” (Defence Leadership Centre 2004), are likely to cause problems when handling National Security Crises which require co-ordination between a complex array of multi-organizational agencies. The “dark side” of charismatic leadership (Howell and Avolio 1992 p.43) is also worth consideration; whilst not suggesting that all charismatic leaders have malevolent intentions, they may take actions which support their own positions rather than those of the organisations they lead. Indeed, it is questionable whether the imposition of any individual’s will upon a group of followers is leadership or something more coercive (Grint 2005 p. 1472). The notion that charismatic leaders require a crisis to be effective suggests they may seek to create one or, using Grint’s (ibid) typology, categorise the situation as Critical in order to use this style of leadership.

The suggestion that any leader, charismatic or otherwise, has the cognitive ability to assess all facets of a situation from all perspectives, correctly assess the nature of the problem and act accordingly, seems questionable as there are a host of psychological distortions in information processing which result in the decision-making process of individuals being flawed (March 1994). For example, Simon’s (1945) concept of ‘limited rationality’ suggests that, as humans have limited cognitive capacity and the analysis process is too long, generally only a limited number of options are envisaged which result in the most viable, as opposed to optimal, being selected (ibid). Indeed, as Janis (1982 p.260) comments, “only the most authoritarian of leaders fails to recognise the peril in relying solely on his own deliberations.” It has, therefore, been seen that whilst charismatic leadership may be useful in swiftly responding to incidents which have been categorised as Critical, this style of leadership has limitations in dealing with complex, multidimensional National Security Crises which, coupled with the constraints of individuals’ cognitive ability, mean that its utility is limited. As Janis (ibid) points out, the usual way to try to counteract the limitations of individuals’ mental functioning is to relegate important decisions to groups. Therefore, self-reliant teams will be examined next to ascertain whether they will be more effective.

In sharp contrast to followers of a charismatic leader, members of self-reliant teams are less dependent on a hierarchical leader, and more empowered to take on leadership responsibilities and authorities themselves (Fisher 2000). Whilst the term self-reliant implies leaderless autonomous groups free from external control and constraint, in reality boundaries exist. The team primarily has responsibility for day to day administration and schedules, whilst the more ‘visionary’ aspects of leadership are still undertaken by someone in a formal position (Jackson and Parry 2008). Team leadership can reduce pressure on the leader to produce all the answers, allowing them to focus more on maintaining effective relationships in the group and enhancing group cohesiveness (ibid). This could be important considering the multitude of actors involved in
National Security Crises, as interferences with sound thinking can occur when members have no sense of loyalty to the group, and regard themselves merely as representatives of different departments with clashing interests and ontological viewpoints (Janis 1982). Although much of the research on self-reliant teams was carried out in manufacturing industry (op cit), and its direct applicability to National Security Crises is unclear, the perceived benefits of self-reliant teams in unleashing creativity and innovation through empowerment and creating flexibility and cohesion across departments suggests that they may have some utility in managing Tame problems. Indeed, Lewis (2006) believes that a team approach to crisis management is best. However, as Janis (1982 p.3) acknowledges, “groups can bring out the worst as well as the best in man”; some of the potential downsides of using self-reliant teams will now be explored.

The cohesiveness encouraged in self-reliant teams, together with their potential isolation from others, could prove counter-productive if ‘groupthink’ sets in. Defined by Janis (1982 p.9) as a “mode of thinking that people engage in when they are deeply involved in a cohesive in-group, when the members’ striving for unanimity override their motivation to realistically appraise alternative courses of action,” ‘groupthink’ is more likely to arise in crisis situations when members work closely together, share the same values and are isolated from accessing expert information (ibid). The effects of ‘groupthink’ can include over-optimism, lack of vigilance, conformity to the group which prevents members from raising controversial issues and questioning weak arguments, the repression of individuals who dissent from the collective view, and a shift towards riskier courses of action than individual members would have otherwise been prepared to take (ibid). Whilst recognising that not all self-reliant teams will be susceptible to ‘groupthink,’ and that its “most corrosive effects” (ibid p.275) can be counteracted by altering the conditions before a crisis emerges, it would appear that the propensity for self-reliant teams to be affected by ‘groupthink’ could make it unsuitable for effectively handling certain National Security Crisis situations.

In addition to ‘groupthink’, problems including lack of accountability and group displacement of responsibility can arise with groups such as self-reliant teams (Grint 2008). Research also suggests that, rather than creating an environment which encourages innovation, teams who have been allowed to form their own norms, rules and expectations are actually more constrained, as the structures have been developed internally and not imposed from above (Jackson and Parry 2008). Group decision making can also be a prolonged process as it encourages lengthy debate. It would, therefore, be inappropriate for the handling of a situation deemed to be Critical, but potentially more appropriate for Tame or Wicked situations. However, in spite of all these potential issues, as Janis (1982 p.12) points out, a group whose members have properly defined roles, with traditions and standard operating procedures that facilitate critical inquiry, is probably capable of making better decisions than any individual who works on the problem alone. Attention will now turn to examine the role of well trained officers.

The potential for using well trained officers to handle National Security Crises is reflected in Weick and Sutcliffe’s (2007) research on High Reliability Organisations (HROs) and their commitment to resilience. Defined as “the intrinsic ability of an organisation to maintain or regain a dynamically stable state, which allows it to continue operations after a major mishap and/or in the presence of a continuous stress” (ibid p. 13), resilience is based upon effective training to ensure deep knowledge of systems, co-workers and individuals (ibid). Scenario-based training provides the opportunity to rehearse actions so that when a crisis occurs, officers involved are mentally prepared and know exactly what to do. The logic of planning for events that are reasonably certain to occur suggests the use of well trained officers may be appropriate for Tame situations. However, as Weick and Sutcliffe (ibid p.23) point out, crises envisioned seldom resemble those that actually unfold and, whilst the expectations created through scenario based training can create orderliness and predictability, they also influence perception and can create blindspots. In addition, as officers may be trained to handle crises in a certain manner, they may actively seek evidence which confirms their expectations and avoid evidence which disconfirms them (ibid). Clearly, whilst the
need for flexibility may be incorporated, the likelihood is that any training assumes that the crisis has already been envisaged and there is a best way to handle it. Thus, whilst well trained officers may play a significant role in the handling of Tame situations, they are not necessarily equipped to handle novel Wicked ones.

Weick (2001 p.92) proposes that the ambiguity and uncertainty surrounding situations categorised as Wicked requires a different leadership style; instead of focusing on decision-making, the leader should focus on ‘sense-making’ instead. Thus the real skill is not in removing uncertainty but in managing to stay effective in spite of it (Grint 2008). ‘Sense-making’ is viewed by Weick (ibid p.93) as “navigating by means of a compass rather than a map”; whereas maps are useful in known worlds which have been charted before, compasses are required when you are unsure of your location and can only provide a general sense of direction. By admitting to their followers that they don’t know the answer, and not presuming there are any generic right answers, Weick (ibid) contends that leaders can help others make sense of the situation by encouraging improvisation and channelling decisions to those with the best knowledge of the matter, such as executives or lower grade individuals. In addition, as there is no pretence regarding the omniscience or expert authority of the leader, and an acceptance that logic and rationality may be inappropriate in an unpredictable situation, followers are encouraged to freely interact to collectively make sense of the situation (ibid). Grint (2008 p. 13) concurs that engaging with the collective and encouraging ‘Constructive Dissent’ is essential for individual leaders trying to come to terms with Wicked situations which are not susceptible to individual resolution. Whilst leaders might traditionally feel inclined to postpone action until they can make the right decision which they then feel compelled to defend, with ‘sense-making’ leaders provide the team with a plausible direction for an indefinite period which is dynamic, open to revision, self-correcting and responsive (Weick op cit p. 97). In a similar vein, Grint (op cit) believes that leaders need to adopt the skills of the ‘bricoleur’ to pragmatically engage whatever comes to hand to formulate ‘Clumsy Solutions’ to Wicked problems. It would, therefore, appear that leaders as ‘sense-makers’ or ‘bricoleurs’ may be the most appropriate to handle complex National Security Crises.

Finally, whilst alluded to throughout, the crucial contributory role of executives and lower grade individuals in the effective handling of National Security Crises, no matter how they are categorised, needs to be highlighted. Under charismatic leadership, the role of executives and lower grade individuals is to validate charisma in their leader and to follow their Command without question (Bryman 1992, House 1976). In a self-reliant team, they are able to take on some of the leadership responsibilities themselves and, if they are aware of the potential for ‘groupthink’ (Janis 1982), can be involved in potential mitigation. This could occur in three main ways; firstly, executives could lead different subgroups which would come together to debate differences; secondly, each member of the group could discuss deliberations periodically with trusted associates and report back reactions; thirdly, outside experts or qualified colleagues within the organisation could be invited to challenge the views of the core members (ibid). The latter two points involve consultations with Subject Matter Experts who are likely to be found throughout the organisation, both at executive level and lower grades. Finally, in crises handled by well trained officers, their roles will be well defined and regularly practiced. The importance of deference to the expertise of executives and lower grade individuals in Wicked situations is highlighted by Weick and Sutcliffe’s (2007) study of HRO’s; decision making is pushed down to the front line with authority migrating to the people with the most expertise, regardless of rank. This pattern is recognised in Weick’s (2001 p.93) hypothesis that the unknowability and unpredictability of crisis situations in the 21st Century will lead to decisions being migrated to those with the expertise to handle them, with less convergence of decision on people entitled by rank to make them.

In conclusion, the way in which a National Security Crisis is constituted in terms of being a Critical, Tame or Wicked situation has implications for the way in which it is handled. Whilst a Critical situation might call for Command associated with charismatic leadership, the unquestioning obedience required of followers, coupled with the flawed notion that an individual has sufficient
cognitive ability to comprehend and effectively handle the situation, suggests that its effectiveness is seriously limited. Conversely, the flexibility and cohesion afforded by the relative freedom from hierarchical leadership enjoyed by self-reliant teams has clear utility in a National Security environment with its multitude of actors. However, the potential emergence of ‘groupthink’ in highly cohesive groups together with the potential displacement of responsibility highlights the weaknesses of using self-reliant teams to deal with complex situations. Both self-reliant teams and well trained officers may be appropriately used in the handling of crises categorised as Tame, where training can prepare for anticipated scenarios which may have happened previously. However, none of these roles can effectively handle a Wicked problem whose novelty and complexity requires a leader in the role of ‘sense-maker’. The analysis has also highlighted the crucial role played by executives and lower grade individuals as followers, team members and, more importantly, custodians of expertise. The significance of their role should not be underestimated for, as Grint (2008 p.21) comments, whilst typically success and failure in handling crises is attributed to individual leaders, more often it is a consequence of collective intelligence not individual genius.

References
The strategic dynamics of international collaboration for defence procurement and support

Group Captain S D Ellard Royal Air Force

Background

The ability of a state to produce armaments was historically a national capability that could deliver a competitive (or combat) advantage over potential adversaries. However, the increase in complexity and cost of modern weapon systems in the second half of the 20th century led many nations to seek collaborative development and production as a potential means of sharing technology, reducing risk and lowering costs. Initially, this was mainly in the domain of large, complex weapons systems such as aircraft and ships, but more recently collaboration has broadened to less complex systems and sub-systems. Nor has the idea been bounded to military applications: starting with Concorde in the 1950s and 1960s, civil aircraft collaboration, especially in Europe, has developed into successful multinational companies such as Airbus, formed by industrial restructuring for the better delivery of joint programmes. Nevertheless, in the military domain, the benefits delivered by multination collaboration for military procurement have not always lived up to their promises.

A recurring theme of the many reviews on defence procurement has been an almost unanimous belief that international collaboration is the only viable model for procurement of major complex platforms in the future. The subject of international defence procurement collaboration is therefore worthy of analysis to understand fully the strategic dynamics of these programmes, the driving factors, the national interests at stake and the manner in which they are delivered. It will be seen that the interests served are varied, not only between the participating nations, but also between national actors and between different levels of strategic perspective.

This paper therefore explores in turn: the UK Ministry of Defence (MOD) perspective; the UK Cross-Government perspective; and the perspective of international organisations such as the EU and NATO, in order to understand the barriers to delivering the benefits that international collaborative programmes promise, and the strategic dynamics behind performance to date. Tools used to achieve these aims include analysis of ends, ways and means, opportunities and threats, as well as concepts such as strategic fit, strategic narratives, the application of power, leadership and the ability to turn strategy into action.

The paper has a bias towards major air platforms, as these systems have been the most vexed area in the international collaborative environment, and also, due to the author’s background experience; broadening this approach to other environments and sub-systems may be a topic for further research.

The UK MOD Defence perspective

There are many varied definitions and descriptions of strategy, but that principally taught by British Military Doctrine is that strategy’s ‘purpose is to balance the ways and means required to achieve stipulated ends, conditioned by the environment and prospective opponents’¹. In this section, ends, ways and means are analysed, initially at the strategic environment of the UK MOD level, in order to demonstrate the importance of international collaborative procurement and support solutions to the UK MOD, and illustrate why the principle is so appealing.
Starting with ‘ends’, the UK MOD is required to equip the UK Armed Forces with affordable equipment and support solutions, and ensure that the right levels of sustainable materiel are available to meet Defence needs. The 2010 UK Defence Review defined the requirement and formed the basis of a plan to achieve balance in these areas. However, global security developments since 2010 have by no means reduced the military requirement and the pressure for more capable equipment is likely to increase. Further pressure may result from the US pivot to the East (possibly leading to UK interests that do not match those of the USA and requiring interventions that do not include the USA), an increasingly assertive Russia, and continued upheaval as a result of the Arab Awakening. These developments suggest that there is going to be increased pressure for ‘more’, both in volume and capability terms.

Turning to ‘means’, a PESTLE analysis (Political, Economic, Social, Technical, Legal and Environmental) of the UK’s strategic environment is dominated by two major influences: Economic and Political. Firstly, the UK economy will require continued fiscal austerity for many years to come and downward pressure on government expenditure is likely to remain. Secondly, there will be greater political pressure to spend on popular, vote-winning departments such as education and health, rather than on defence. Both of these factors indicate that there will be ‘less’ means to deliver Defence’s needs in the near future.

Finally, ‘ways’. The dominant theme for strategy formulation in the current defence environment is to find ways to make defence expenditure cheaper so that ‘more for less’ can be achieved. A large proportion of the defence budget is spent on equipment, so there is a demand for an alternative means of delivering equipment procurement and support in a more affordable manner that offers greater value for money. This is not a new ambition. Indeed, it has been the subject of a large number of reviews, including The Gray Report and Enabling Acquisition Change (within the last 10 years alone). Still, almost all of these reports claim that the fiscal gap can be filled, at least in part, by delivering the perceived benefits of international defence procurement.

Strategic analysis therefore presents a highly compelling case for the promotion of international defence projects, wherever possible, in order to achieve strategic balance between ends, ways and means. However, experience has shown that this conclusion is deceptively simple and naïve, and the reality is far more complex. The following sections offer an analysis of the opportunities and risks presented by international collaboration to determine if the seemingly simple and attractive strategic principle of international collaboration can meet such demanding requirements.

When considering the opportunities, the principal benefit offered by international collaboration for defence procurement is reduced overall costs per nation. Advocates argue that this approach offers increased production runs and, hence, lower fixed costs per unit item (i.e. economies of scale). Furthermore, they claim, large numbers of nations collaborating on a project reduces risk, by sharing risk among a larger number of contributors, and shares the technological efforts of a broader range of facilities, thus establishing a more reliable, and effective capability development path. Commercially, there is also increased scope for better value for money as the base from which suppliers can be chosen is made wider and the ‘best athlete’ for the work can be selected. In addition, large numbers of nations acting together avoids the prospect of alternative products being developed and produced in competition, thus saving unnecessary expenditure, energy and resources. It is also claimed that international collaboration offers the opportunity for a nation to take part in a defence procurement project that, otherwise, would have exceeded its industrial and/or technological capacity. Nevertheless, it has long been understood that the added complexity due to the number of nations involved in an international project dilutes the benefits. Additional industrial interfaces result in a non-linear relationship between the number of nations and overall savings achieved. Indeed, it has been postulated that a ‘square root’ relationship exists. For example, if the ‘normal’ cost of a programme development is £X (i.e. the cost if a nation
were to develop the programme alone) and n nations are participating, then the increased total cost is \( X \sqrt{n} \). Therefore each nation will contribute the square root of \( n \) divided by \( n \):

\[
\text{Cost per nation} = \frac{X \sqrt{n}}{n}
\]

So, for example, if 2 nations are collaborating, then each would pay \( \sqrt{2} \), i.e. 71\% of ‘normal costs’ (even though the total is 142\% of ‘normal’ costs). Similarly, if 4 nations were to collaborate, then each would pay \( \sqrt{4} \), i.e. 50\% of normal costs (even though the total is 200\% of ‘normal’ costs). Thus costs per nation, if this model is correct, will fall per nation as the number of participating nations increases; however the net benefit per nation of adding an additional partner decreases as the number of nations gets larger (for example, there is only an additional saving of 1\% per country if the number of nations is increased from 10 to 11). Therefore, once the major savings have been achieved by collaborating with only a few nations, there seems to be little financial benefit in increasing the number of partners further.

Threats nevertheless exist. The first is a consequence of nations seeking to win domestic industrial work share, often in proportion to the amount of equipment being procured (the principle of ‘juste retour’). As a result, segments of industry are not chosen on the basis of the ‘best athlete’ principle, where the most competitive, reliable and experienced company receives a contract, but allocated instead on the basis of work-share to a company that may have no experience or background in that area of expertise. Indeed, the company may have been chosen specifically to develop an industrial capability that hitherto had not existed in that nation. This can result in increased costs, unreliable and/or underperforming equipment and delays. A further threat to the delivery of international collaboration benefits, often quoted by detractors, is the inability of nations to agree on a common capability requirement, and it is claimed that the consequential price of ‘being different’ vastly increases costs. This results either in a base version that is the lowest common denominator in capability terms, as nations will refuse to pay for capabilities they do not need, or with a price premium for those who wish to deviate from the baseline, as some nations will find it intolerable to be deficient of the capability. Furthermore, while viewed by some as being an opportunity, others view a single-sourced solution as a threat, as it does not generate the creative tension and innovative thinking that competition can stimulate, unless alternative means for driving innovation, such as incentivised gain-share or recycling, are adopted. It is also possible that contrasting strategic cultures may lead to misunderstanding, misinterpretation and, thus, further delays and complications, especially if officials and contractors assume that negotiating styles and contractual designs are done in all nations in the same manner as at home.

While these threats certainly contribute to increased costs and under performance, the greatest single risk to the delivery of collaborative procurement programmes to performance, cost and time is more structural. Each nation has its own rules, regulations, processes and procedures for: procurement; seeking approval to proceed with stages of a programme (both internal and political); placing commercial contracts (sometimes the favoured approach for one nation is illegal and not permitted in others); accounting treatments (for example the distinction between capital investment and running costs); safety (and airworthiness) clearances; and evidence, documentation and authorisations. The cumulative impact of this diversity and complexity can lead to decision making gridlock and incurs considerable delays and therefore increases cost. The cost of delays caused by this incompatibility is not simply the additional cost in producing the equipment. Further costs are incurred indirectly when new equipment is procured to replace existing equipment, as delay means that additional costs are incurred maintaining and operating expensive legacy equipment at the same time. The financial effects of delay are effectively increased twofold. There are also fears about control of supply becoming a threat when entering
collaborative defence procurement agreements. Collaboration brings with it a dependency on other nations’ industries and governments to provide continued equipment support, thus risking autonomy on operations, freedom of independent manoeuvre and a degree of sovereignty. In addition, there is a danger that the UK’s reputation in international collaborations jeopardises future endeavours, as the UK’s track record in meeting pledges for international collaboration is not good. ‘Perfidious Albion’ has a habit of making promises to collaborate, and then breaking them if a cheaper alternative presents itself. If this trend continues, the UK may lose the trust of partner nations and find itself unable to lead or participate in future programmes with the same degree of influence.

Strategic ‘fit’ is achieved if a strategy is consistent with its environment, with its resources and capabilities, and with its organization and systems. It is a useful test for the effectiveness of a strategy, beyond ends, ways and means, because it can explain why a strategy can be successful in some environments, yet fail in others. The description of opportunities and threats above demonstrates that international collaboration will not achieve strategic ‘fit’ unless frameworks are in place to ensure alignment of capabilities and procurement procedures. Indeed, the failure of some international programmes may be due to the fact that they were entered into without a thorough understanding of the complexities involved and represents a failure to achieve strategic ‘fit’ due to incompatibility of procurement processes and procedures between partner nations. If strategic fit cannot be achieved then the outcome will be the same as for many international programmes: procurement collaborations that run at the pace of the slowest, deliver capability that is either the lowest common denominator or an expensive bespoke product, and are over budget. Therefore, the greatest consideration to be made before embarking on an international defence procurement collaboration is to understand the differences and the potential incompatibilities of the respective procurement, commercial and financial systems of the participating nations. Scrutiny needs to be broadened beyond capabilities and operating procedures; in addition to seeking partner nations that have the same operational requirement and equipment needs as the UK MOD, the choice should also be influenced by the compatibility of their systems and processes.

Given the imperative to deliver ‘more for less’, it is perhaps puzzling that in-service support does not feature more prominently in the context of international collaboration. Major defence platforms have in-service lifespans of several decades and their complexity and safety criticality mean that in-service support, maintenance and upgrades often cost far more than initial procurement costs. The initial strategic analysis that demanded ‘more for less’ could therefore yield greater benefits if the same attention was applied to support, and is an opportunity to be further exploited. Indeed, the benefits of international collaboration seem to be valid across the entire CADMID (Concept, Assessment, Demonstration, Manufacture, In-Service, Disposal) cycle. In spite of this, it seems that procurement is inherently more political than support and attracts greater attention. It may also be the case that, in many countries, procurement and support are administered by separate organisations and, unlike the UK, the trade-off between procurement and support cannot be made as easily. Indeed, in the UK, the distinction between procurement and support is becoming increasingly blurred and it is in this area that the UK has made the most innovative changes in recent years, with ‘contracting for availability’ and ‘partnered support’ saving huge proportions of support expenditure. As these changes continue to be developed, especially taking into account the potential offered by Lord Levene’s Defence Reform report, then UK procurement and support procedures are likely to become even more divergent from potential collaborative partner nations. As this divergence has been shown to be the greatest threat to successful collaboration, this would seem to paint a gloomy picture for the future prospects for such ventures.
The weaknesses revealed by earlier disappointing international collaborative programmes, especially those managed by the NATO Eurofighter and Tornado Management Agency (NETMA), led European Nations to devise alternative mechanisms to avoid repeating the failings that had been exposed. As a result, an organisation called OCCAR (Organisation Conjointe de Coopération en Matière d’Armement - the Organisation for Joint Armaments Co-operation) was formed to overcome these vulnerabilities. Its mission is:

“To facilitate and manage collaborative European armament Programmes and Technology Demonstrator Programmes, through their lifecycle to the satisfaction of our customers.”

As the head of OCCAR explained in his address to the UK Defence Forum in February 2014:

‘OCCAR was initially formed by France, Germany, the UK and Italy in the 1990s and achieved legal status in 2001. By 2005 Belgium and Spain had joined also, making up the current 6 Member States. In creating OCCAR, its founding nations were determined to solve the major problems that had damaged previous cooperative programmes. I believe that they achieved that, and today’s successful OCCAR programmes owe a great deal to the courage and vision of the people who founded the organisation back in the 1990s.’

The head of OCCAR goes on to claim that:

‘The key is to establish a programme properly, with clarity over the intent, a clear delivery plan, a sound understanding of the risks and uncertainties, and contingency plans and resources to match. I believe that the OCCAR model embodies those needs, and also the solutions to avoid many of the problems that beset earlier cooperative Defence programmes. It provides a highly evolved and effective system with the flexibility … to include non-OCCAR Member states as participants on equal terms.’

If this is the case, then OCCAR would seem to offer a potential means to achieve ‘strategic fit’ despite otherwise incompatible procurement, financial and commercial systems and creates a strategic environment that allows international collaborative defence programmes to succeed. An analysis of the actual benefits delivered by OCCAR, in comparison to traditional collaborative models and bilateral arrangements, would therefore be another area of useful further study.

A strategic narrative is the ‘story’ or description of a course of action that makes it a compelling strategy to follow for a chosen audience. However, within the UK, there is no defined strategic narrative for international collaboration for defence equipment procurement. There are traces of a narrative in studies such as the Gray Report, the Defence Industrial Strategy, and the Defence Growth Partnership, but nothing consistent, nor claimed by any owner. Indeed, the author has had to go as far back as the 1998 Defence Review to find anything that comes close to resembling a strategic narrative where it is stated that international collaboration:

‘offers tangible military, economic and industrial benefits and it is essential that the UK remains at the forefront of developments in this area of joint endeavour.’

Perhaps the lack of a formal strategic narrative to support international defence equipment collaboration is because the topic is viewed as being too simple and self-evident? From a spectrum of sources, RCDS has heard: ‘it’s the way ahead; ‘we can’t afford any other way; ‘we can’t do it by ourselves’ etc. Indeed, the strategic analysis above, highlighting the very pressing need for defence equipment procurement and support to be cheaper, explains why a concept such as international collaboration is so enticing, and the informal, unwritten strategic narrative is extremely attractive. It seems to offer everything one could desire: lower risk, lower costs, reduced
risk, and access to superior technology and quicker delivery of the same capability as other nations.

The simplicity of this informal, unwritten narrative is, nonetheless, deceptive and misleading, giving the impression that one simply needs to find a willing partner, sit back and enjoy watching the benefits flow in. Alas, the track record of many projects shows this not to be the case. Early examples such as the Anglo-French Concorde, as well as the more recent 4-nation Eurofighter Typhoon shows that the adoption of an international collaborative model is by no means a guarantee of success. Getting the narrative right is therefore very important, otherwise it will seduce actors into believing that it is easy. They will be inclined to a psychological preference for intuitive decision making, rather than the rational decision making needed. Because the informal, unwritten strategic narrative as it stands is so misleading, a number of natural follow-on questions result: Who should own and shape the narrative? What should the narrative be? And how should the narrative be communicated to best strategic effect? Within UK defence, it can only be Defence Equipment and Support (DE&S) who should own the narrative. And, clearly, the narrative needs to be more precise than ‘buy internationally: it’s cheaper!’ Perhaps the following narrative would be more appropriate:

‘International collaboration offers the potential for military, economic and industrial benefits, including a significant reduction of technical risk and costs throughout the entire CADMID cycle. However, care should be taken to ensure compatibility of national processes, possibly adopting similar lines to those employed by OCCAR’.

Furthermore, thought needs to be given to the communication method. The closest DE&S has to a corporate memory, where lessons learned and doctrine on how activity should be conducted is recorded, is the Acquisition Operating Framework (AOF). Yet, there is no guidance in the AOF, other than some specific technical commercial procedures, on how international collaborative procurement should be carried out and no narrative to support why or when it is beneficial. This deficiency is representative of an overall weakness within DE&S on analysing, learning and addressing lessons identified and recording experience within formal doctrine. This doctrine need not be so prescriptive that it stifles innovation, but sufficient to avoid repeating the errors made by earlier generations.

Earlier analysis applied the UK MOD’s ‘ends, ways and means’ approach to strategy. An alternative perspective offered by Lawrence Freedman defines strategy as being: ‘about getting more out of a situation than the starting balance of power would suggest’. Freedman implies that the wise application of the power available can deliver a more highly advantageous outcome by utilising a superior strategy. In the context of the UK’s approach to international collaborative defence procurement projects, it invites questions such as: ‘How much power does the UK have?’; ‘What is the source of UK power?’; ‘Is this power being used to best effect?’ Tangible power, in terms of negotiating influence over performance, cost and time, comes primarily from the proportion of equipment to be procured in relation to other nations; a majority take off from the production line should ‘buy’ a proportionate amount of influence. In programmes where the UK has the lead and the greatest proportion of industrial offtake, like Eurofighter for example, then the UK should have the greatest power. However, in this case, power has been diluted as the management agency responsible for the development, production and support of the aircraft operates on a strict consensus basis, where all partner nations have equal influence over decisions. This makes the minority stakeholder, Spain, far more powerful than its starting balance of power would suggest, and the UK less so. In contrast, the UK is a minority partner to the USA in the Lightning II, F-35 programme. Here, the UK’s reputation for understanding capability and the employment of airpower, as well as innovative techniques developed in the field of procurement and support, gives the UK more influence than the proportion of aircraft being delivered would otherwise
suggest. A further source of UK power in the international arena is the advantage of being native English speakers. English is frequently used as a common language within international programmes, and the ability to articulate ideas clearly and persuasively in the English language can give a powerful advantage to those with the strongest skills. Therefore, the UK should be seeking to maximise the reach of this powerful advantage by placing UK officers in leadership roles within international procurement and support management agencies so that this advantage is fully exploited.

While the preceding analysis illustrates that it remains potentially viable to derive benefit from collaboration, the identification of a winning strategy is, in itself, insufficient to deliver the outcomes needed by the UK MOD. The strategy needs to be turned into action. In his book ‘The Art of Action’, Bungay describes three gaps that prevent the effective implementation of strategy: the effects gap, the knowledge gap, and the alignment gap.

The two most dominant gaps affecting the procurement and support strategy at the UK MOD level are the effects gap and the alignment gap. UK does not suffer from a knowledge gap; defence procurement is very highly scrutinised by both internal and external bodies, and performance management processes are widespread and have access to volumes of collected data. Yet, disappointingly, experience shows that there is clear difference between what is hoped to be achieved by international collaboration and what is actually achieved: an effects gap. This is closely related to the threats and weaknesses in strategic fit, and measures such as establishing compatible procurement systems and realising the benefits claimed by supporters of an OCCAR-like structure will go a long way towards closing this gap. Caution must be applied here though, as Bungay notes that a traditional corporate response to this gap is to apply more detailed controls. He argues that this is counterproductive and that instead, within strictly defined boundaries, individuals should be encouraged to use their initiative to fulfil strategic intent and retain freedom of decisions and actions.
There is also a gap between what we want people to do and what they actually do: an alignment gap. This is closely related to the application of power. While the UK MOD may wish other nations to follow strategies that are aligned to UK MOD’s performance cost and time targets, the UK cannot compel other nations to follow this course of action, as nations may well be following their own national interests and be pursuing different goals. It is in this area that the UK will need to use its levers of power, listed above, to persuade and guide partner nations to follow actions that best serve the national interest. This principle can also be applied to the individual level working within organisations, such as DE&S. Enhanced performance management measures and enabling effective measures of motivation and the introduction of incentives that match departmental goals (the ‘pyramid of purpose’) can generate behaviour to close the alignment gap. This is the type of change that the current Chief of Defence Materiel is attempting to embody in DE&S with his Materiel Strategy.

**The national cross government perspective**

The first section of this paper explored measures that could realise the potential offered by international collaborative defence procurement at the UK MOD level. However, the same exercise at the UK cross-governmental level paints a distinctly different picture. At this level, the Grand Strategic level, the strategic environment is concerned with more than simply the performance, cost and time targets set by the UK MOD, and has to consider broader interests such as employment, industrial sustainment, technological development, overseas commitments, regions and countries of strategic relationships, the security of overseas trade, and access to critical resources. The same style of analysis is now carried out at this level to identify how the strategic perspective is altered.

In this context, international defence procurement collaboration may be a ‘way’ to meet an ‘end’ that comprises far broader aims than those followed by Defence alone; indeed Defence’s performance, cost and time targets may need to be subordinated in order to achieve this end. The UK national, cross-governmental grand strategic perspective reveals a different set of opportunities and threats, implying the need to adopt a different strategy, and one in which the UK MOD may have to play a different role to that which it would otherwise follow if solely pursuing internal departmental goals.

At the cross-governmental level, international collaboration on defence procurement programmes offers a broad spectrum of rich opportunities, in addition to those for the UK MOD described in the previous section. The Defence sector is an important employer, and joining an international collaborative defence programme can secure employment across the entire nation, not only at major assembly prime contractor level, but also cascaded down through sub-contractors and suppliers. These are high quality, highly skilled jobs, and the associated technology can stimulate research and innovation that has secondary benefits for the entire economy. Partnership in these fields can also have great political and diplomatic significance and can generate influence with strategic partner nations, a good example being the UK-France Partnership signed at Lancaster House in 2010. Furthermore, the prospect of trade and export offers further benefits, especially in regions and countries where the UK wishes to increase influence and strengthen relations. Indeed, the role of export plays a distinct and unique role and the distinction between international collaboration and defence export is becoming blurred. Viewed purely from the UK MOD level, the role of export could, at first glance, be seen as a secondary activity, adjacent to the main business of procurement and support. Yet at the UK cross-government level, success with export campaigns can deliver many of the same benefits that ‘traditional’ international collaboration offers during early development and production phases: assembly lines are increased and extended, and greater benefits of scale are achieved. Indeed, as export customers increasingly demand industrial participation as part of the negotiations, they can be viewed in the same light as the original
partner nations; they are simply entering the programme at a later stage of development. Those joining at the start are ‘early adopters’ and enjoy greater influence over design and work share, while later export customers are ‘late adopters’ with less influence over design, yet are able to avoid the risks associated with the start of the programme. Furthermore, export customers may be able to invest in upgrade capabilities and thus make spiral development more affordable for the original partner nations.

In addition to the threats posed at the UK MOD level, the greatest hazard to the adoption of a viable grand strategy is the sheer diversity of demands from the plethora of stakeholders, each of whom is seeking different goals from international collaboration. The relatively simple, straightforward UK MOD aims of achieving the right levels of capability in the right time for minimum costs become drowned out by competing strategies. There are those who wish to gain influence and improved performance, cost and time by advocating Commercial (or Military) off The Shelf (COTS/MOTS) solutions, and thereby achieve stronger alliances with those able to provide equipment such as the USA. There are those who advocate a Defence Industrial Strategy that makes clear which segments of the Defence industry are viewed as vital ground for the UK, in order to maintain employment and sustain industrial capabilities. There are also those who advocate exporting equipment to those countries where the UK wishes to ensure security and stability, increase revenue and gain political favour, and strengthen bonds with regions deemed to be strategically significant. These strategies appear divergent, incompatible and seem to have been created in isolation. They are certainly incoherent and risk failure unless a strategy that identifies clear national interest can be developed. Furthermore, even if such a strategy can be developed, the combative process to reach a solution can involve a huge amount of energy and debate expended among competing parties, and can induce delays that dilute the strategic effect and lead to increased costs.

Despite the challenges and threats posed when implementing international collaboration at this grand strategic level, the potential rewards and opportunities are so compelling that they are worthwhile pursuing, and strategic fit can be realised with the right degree of in-depth, strategic thinking. A hypothetical example illustrates this point effectively: say an export customer exists for a high-end weapons system platform employed by UK Armed Forces. The Foreign Office is keen to develop links with this country, as it is of future strategic importance. The UK Armed Forces would like to fit an additional sensor to the system, but this option is currently unaffordable for the MoD. The export customer would like to fit the sensor too, but will only do so if it has the reassurance that the UK Armed Forces have already successfully integrated the sensor. The sensor provides a large number of jobs for UK Industry, and also stimulates research and intellectual capital in an area that offers long-term potential for Industry. Yet the situation would remain at gridlock due to the absence of any funding to prime the solution. If, say, the Department of Business, Innovation and Skills were to fund the research and development, then the system would be affordable for the UK Armed Forces and Industry would benefit from the research and development and work involved. The export customer would then proceed with the purchase of the platform (and sensor) that would bring funding to industry, secure jobs, and bring additional revenue for the government. Therefore, by taking a cross-government approach, all parties can achieve a positive outcome, but it requires government departments to collaborate in a manner that is not incentivized by normal departmental strategies. A cross-government, co-ordinated approach is the only manner in which such a mutually supporting arrangement could be developed; without central co-ordination, and relying on bilateral arrangements and ad-hoc arrangements, the process would either be extremely protracted, or would fail to materialize.

The author has been unable to find any documented narrative to describe the strategy of collaborative defence procurement at the national cross-governmental level, and it is by no means clear who should own, develop and communicate the strategic narrative; few cross-departmental
bodies come to mind that span the right breadth of influence, with the exception of the National Security Council (NSC). If a narrative were to exist, it should probably develop further the UK MOD narrative to include references to the broader aims of collaboration and the necessary co-ordinating function needed. Perhaps, something along the lines of the following:

‘International collaboration offers the prospect of political, diplomatic, employment, military, economic, technological and industrial benefits, across many government departments, as well as the potential to reduce technical risk and costs throughout the entire CADMID cycle. However, cross-governmental co-ordination to develop the strategic aims of the international collaboration is vital in the planning phase, and for the delivery phase care should be taken to ensure compatibility of national processes, possibly adopting similar lines to those employed by OCCAR.’

It has been shown that to align potentially conflicting strategies among different government departments and create a viable ‘grand strategy’, a coordinating body is needed to avoid competition and conflict. The most suitable body would seem to be the NSC, as it is only at this level that departments (with naturally differing agendas) meet to develop national strategy, although only the most significant programmes would merit the attention of such a senior meeting. In response, the NSC has set up the Export Ministerial Working Group (EMWG - a formal subcommittee of the NSC) to allow it to debate these issues regularly.32 This is an extremely positive development that offers a template for future cross-governmental co-operation. This can only be successful with strong leadership, most likely requiring personal direction or intervention by the Prime Minister, to be able to drive and direct a coherent cross-governmental agenda. A further measure that offers potential, again benefiting from central leadership and direction from the Prime Minister, is the cross government initiative named ‘The Defence Growth Partnership’. This has as its vision:

*The Defence Growth Partnership will secure a thriving UK Defence Sector delivering security, growth and prosperity for our nation*33.

Its strategy is to:

*take a fresh and ambitious approach through a joint commitment to work together to develop new opportunities by building on our nation's strengths and deliver growth through innovative and tailored solutions*34.

This is not only a cross-governmental initiative, but includes partnership with a broad spectrum of the UK defence industry. While not aimed at international collaboration in its own right, it does offer the potential to align the competing Departmental and industrial goals in a manner that would allow both international collaboration and export to succeed, despite the threats listed earlier. Success in both endeavours will require not only very senior leadership, but also strong and effective leadership at all levels if a potential ‘permafrost’ layer of management (perhaps those personally not supportive of the grand strategy) is not to thwart grand strategic ambitions. For the UK MOD, this may involve accepting a price premium, a compromise on capability, or a delay to the delivery of a capability – a concept that will be difficult for Defence to accept if the principle of ‘costs fall where they lie’ is adopted, and Defence is required to pay a premium for benefits that will be enjoyed by other departments.

If Bungay’s model for effective strategy implementation35 is revisited at the cross-government ‘grand strategic’ perspective, a different picture emerges to that seen at the departmental level. In particular, it can be seen that the greatest impact at this level is the alignment gap. It has been shown that without some form of a co-ordinating authority, different government departments
will not be aligned. Nonetheless, even if they can be aligned by bodies such as the NSC, then this alignment may not be effectively cascaded down throughout the organisation. For example, unless the UK MOD updates its performance management systems to reflect the goals set at the grand strategic level, then personnel will continue to follow the traditional strategic goals of performance, cost and time. This is especially relevant if bonuses and other motivation levers are linked to legacy targets. Again, Bungay advises caution to avoid the traditional corporate response of issuing more detailed instructions; instead, he proposes that it is more effective to ‘communicate to every department as much of the higher intent as is necessary to achieve the purpose,’a concept familiar to military personnel as ‘mission command’.

The international perspective

It is not only government departments and nations that seek international collaboration for defence equipment, but international organisations too. Both NATO and the EU have identified that the dilemma of increasingly challenging security ‘ends’ and decreasing financial ‘means’ demands ‘ways’ that can deliver ‘more for less’. In response to this challenge, NATO has sought to implement a collaborative defence equipment programme called ‘Smart Defence’, which aims to ‘encourage Allies to cooperate in developing, acquiring and maintaining military capabilities to undertake the Alliance’s essential core tasks agreed in the new NATO strategic concept’. In addition, the EU is promoting a similar programme called ‘Pooling and Sharing’ that ‘comprises a series of actions to support cooperative efforts of EU Member States to develop defence capabilities’.

The principal opportunities offered by the NATO and EU initiatives are similar to those sought by the UK MOD; namely greater overall capability at an affordable price brought about by a reduction in equipment costs for each participating nation. The arguments are the same: increased production runs and lower fixed costs per unit item; shared risk among a larger number of participants; and the avoidance of duplicated effort on unnecessary competitions.

When considering threats, though, NATO’s reputation for delivering procurement programmes is poor, and the NATO management agency for Tornado and Typhoon, NETMA, attracts wide-ranging criticism for the delays and cost overruns associated with these programmes. Weaker still, the EU has no background or experience in delivering international defence programmes at all. RCDS has heard that neither programme has any new ideas that differ from earlier disappointing results, and there is little likelihood that performance will be improved this time. It seems that NATO and the EU have adopted the simplistic strategic narrative of ‘collaborate internationally because it is cheaper’ without understanding the complexity that underlies successful delivery, proposing any solutions to overcome the procedural challenges, or acknowledging the existence of competing national aims. While both organisations have passionate advocates for the respective programmes, there seems to be less enthusiasm from nations themselves, and there is little evidence of a ‘coalition of the willing’ prepared to drive the change programme forward.

The strategic fit therefore seems fragile for both NATO’s Smart Defence and the EU’s Pooling and Sharing. There is balance between ends, ways and means, but the strategy does not fit the strategic environment, the proposals do not fit with nations’ interests, and the vision is not fully shared. The lack of new or innovative ideas to differentiate these initiatives from earlier, discredited attempts is particularly disappointing. At face value, the EU’s ‘Pooling and Sharing’ would seem to have a greater chance of success, as the EU has access to political, economic and industrial levers that NATO does not, and it has been shown that international collaboration is essentially a political, economic and industrial activity. However, despite these capabilities being available to the EU, it is assessed as extremely unlikely that the UK would hand over sovereignty in this area to the EU. Perhaps if the EU were to advocate industrial restructuring and rationalization
of the European Defence Industry, then the environment would be more amenable to this venture?

In contrast to UK MOD and UK cross-governmental strategies, NATO and EU strategic narratives do exist and are well communicated via various channels, including the internet. The weakness is that they are owned and developed by the organisations themselves, and are incoherent with individual national intentions and nations’ interests. The existence of a well communicated narrative in its own right is thus no guarantee of strategic success.

A strategic vision is vital for all change programmes. It can simplify the decision making process; motivate people to take the right action; and help co-ordinate the decisions of many people in diverse locations. The vision needs to be shared by all parties and there is little evidence that this is the case with all member nations. Therefore, while a vision is clearly and succinctly described by both NATO’s and the EU’s strategic narratives, neither has sufficient support from partner nations and neither is truly coherent with each nation’s own vision; each nation is following its own national interests, and is unlikely to subordinate these to a greater NATO or EU vision.

The outlook for the NATO and EU strategies seems equally gloomy when Bungay’s model is applied. While there is no knowledge gap, there are certainly effects and alignment gaps. The effects gap is clear: NATO initiatives have disappointed in the past and nothing new is on offer; therefore, unless innovative approaches are adopted, the same disappointing outcomes will repeat themselves in the future. As for the alignment gap, nations do not appear committed to the vision and are unlikely to subordinate national interests to it. Unlike the power that the UK Prime Minister is able to apply over disparate departments using levers such as the NSC and initiatives such as Defence Growth Partnership, neither the EU nor NATO has the power to compel nations to follow their strategies. Therefore, unless either organisation is able to persuade nations to follow their leadership, the prospects for either initiative do not appear hopeful.

**Conclusion**

The UK, like many other Western nations, seeks to reduce defence procurement and support costs due to budgetary pressures and an increasingly unpredictable defence and security environment. The potential exists for these aims to be met by harnessing the benefits of international collaboration. However, the delivery of these benefits has been elusive as programmes overrun for a variety of reasons, including: non-competitive tendering; misalignment of capabilities; incompatibility of procurement processes; and divergent accounting and commercial regulations. Many of these issues can be addressed by adopting principles similar to those advocated by OCCAR, or an OCCAR-like construct, if the benefits claimed by this organisation can be independently demonstrated. It is, therefore, important for a realistic strategic narrative to be owned, defined and communicated to ensure that the underlying principles and inherent complexity are not misunderstood. The role of power, and its wise employment, can also contribute greatly to the likelihood of strategic success. As a more strategic view is taken across Government departments, complexity increases even further and the ability to craft a coherent strategy and align divergent organisations becomes even more challenging. Indeed, cross-governmental and political influences may demand an international solution that is less than optimum for Defence, while being in the overall national interest. Therefore, strong leadership is required on such occasions to turn ‘strategy into action’, as cross-governmental and political drivers may demand that departmental stove-pipes act in a manner contrary to their internal objectives. The role played by the National Security Council (NSC) and the Defence Growth Partnership will be key here, and could form the basis of a strategic narrative that is currently lacking. Initiatives by international bodies such as NATO (SMART Defence) and the EU (Pooling and Sharing) face real challenges, as they offer little new beyond the attempts that have failed in the
past and align neither with departmental strategic goals nor with national cross-governmental grand strategic goals.

Endnotes

2 MOD Global Strategic Trends – Out to 2040, p. 15.
3 Gray B (2009), Review of Acquisition for the Secretary of State for Defence (MoD) p. 83.
5 Air Marshal S Bollom, Chief of Materiel (Air), in an interview with the author on 25 March 2014.
6 Hartley & Braddon (2012), Collaborative Projects and the Number of Partner Nations, Unpublished, p. 5.
8 Mr T Rowntree, Chief Executive OCCAR, in an interview with the author on 25 March 2014.
9 Benoit Gomis (2011), Franco-British Defence and Security Treaties: Entente While it lasts? Industrial Co-operation: The devil is in the detail (Chatham House) p. 17
11 Air Marshal S Bollom, Chief of Materiel (Air), in an interview with the author on 25 March 2014.
12 Acquisition Operating Framework, version 2.11.33 – September 2013.
14 Mr T Rowntree, Chief Executive OCCAR, in an interview with the author on 25 March 2014.
17 Gray B (2009), Review of Acquisition for the Secretary of State for Defence (MoD). p83.
18 MOD Defence Industrial Strategy (2007),
19 UK Government Defence Growth Partnership (2013),
24 Air Marshal S Bollom, Chief of Materiel (Air), in an interview with the author on 25 March 2014.
28 UK-France Summit 2010 Declaration on Defence and Security Co-operation, Published 2 November 2010.
29 Gray B (2009), Review of Acquisition for the Secretary of State for Defence (MoD), p. 84.
30 Ms S Mason, Director General Exports and Commercial Strategy, in an interview with the author on 31 March 2014.
31 Ms S Mason, Director General Exports and Commercial Strategy, in an interview with the author on 31 March 2014.
32 UK Government Defence Growth Partnership (2013),
33 UK Government Defence Growth Partnership (2013),
34 UK Government Defence Growth Partnership (2013),
40 Air Marshal S Bollom, Chief of Materiel (Air), in an interview with the author on 25 March 2014.
41 Kotter, The 8 Step Change Process, STEP3: Develop a Change Vision
42 Francis Milton, UK MoD, in an interview with the author on 28 March 2014.
Deterring non-state actors: A dialogue of perspectives

LT COL M J D ACKRILL

Introduction

Few will forget where they were and what they were doing on 11\textsuperscript{th} September 2001. The fires at the site of the World Trade Centre burned for one hundred days.\textsuperscript{1} The impact on the theory and practice of international relations continues today, not least in considering how deterrence might stop an event like 9/11 from happening again. This paper will examine whether and how that strategy could be used to counter threats posed by non-state actors in the 21\textsuperscript{st} Century.

Despite the infrequent but high-impact effect that non-state actors can have on national security agendas, there has been little or, at best, localised intellectual effort applied to considering how they could best be deterred. As the then Chairman of the US Joint Chiefs of Staff, Admiral Mullen stated; “deterrence today is tougher and more complex … yet we have done precious little spade-work to advance the theory”.\textsuperscript{2} The Foreword to the UK Defence Plan of 2014 (the MOD’s top level management plan) captures the complexity of the challenge. It recognises that many contemporary security threats defy orthodox definitions. Furthermore, the dispersion of power in the international system means that some of these threats cannot be dealt with by traditional military methods.\textsuperscript{3} While the non-state actor threat may be resurgent, options to influence it are elemental. Cooper captures it rather neatly when he states that; “soldiers and diplomats are, in the end, trying to do the same thing: to change people’s minds. And both are subject to the same problems: not knowing what is in those minds nor how they will react when you try to change them”.\textsuperscript{4} If deterrence is as old as strategic history, so too are its limitations and flaws, perhaps uniquely so because successful deterrence is difficult to prove and, therefore to learn lessons from. The difficulties appear even starker when considering the nature of non-state actors. What levers can be exercised against a challenger without territory, population or recognisable economic apparatus? Can you deter an idea or cause that manifests itself outside of the state construct? These are areas which this paper aspires to address.

‘Deterrence’ is a concept plagued by a wide variety of definitions (perhaps a reason for perceptions of historical success or failure). While it is not unusual for international relations concepts to be contested or misused, with deterrence the implications are more significant than usual. As Gray states; “such abuse really matters because of the misuse of words that may well reflect, and it must encourage, unsound thinking”.\textsuperscript{5} The difference of opinions lies partly with the disputed boundaries of coercion, deterrence and compulsion (or, in the American vernacular, compellance). For the purposes of this paper, Schelling’s categorisation of deterrence and compulsion as sub-sets of coercion\textsuperscript{6} is the baseline. Deterrence is defined as “an institutionalized, coercive social logic aimed at dissuading the use of violence”.\textsuperscript{7} The central and distinguishing aspects are that the deterred actor chooses inaction because of the perceived intolerability of the consequences, and does so before any action has occurred.\textsuperscript{8} Time is a factor and also a challenge for practitioners of deterrence. Those who seek to compel require an adversary to take action within a specified time. For those who seek inaction, the timescale is potentially infinite. In the broadest sense, deterrence has two components. First is the capability to impose a future cost, and the credibility to do so. Second is the ability to accurately communicate the threat of the possible future cost and, also, that inaction is the condition with which the adversary can avoid incurring that cost.\textsuperscript{9} These ingredients apply to deterrence of any type; nuclear or conventional, against state or non-state adversaries.

Having defined what, for the purposes of this paper, deterrence is, it may aid clarity to give examples of what it is not. Some offer ‘self-deterrence’ as a distinct sub-category but, because an actor must choose to be deterred, all deterrence is self-deterrence. Terms such as ‘deterrence by denial’\textsuperscript{10} or ‘offensive deterrence’ are also problematic. Deterrence has inaction as its goal; compulsion applies when the coencer wishes the challenger to stop an action that is already taking place. US airstrikes against Libya in
1986, in response to terrorist attacks against American servicemen in Europe, has been offered as an example of ‘offensive deterrence’\(^{11}\), but would be better described as compulsion. The last example of unhelpful use of the term is when deterrence becomes retroactive, a policy employed in what Adler describes as “a paradoxical response to aggression”.\(^{12}\) Not only is it inaccurate to label Israeli operations against Hizballah in 2006 and Hamas 2009, for example, as deterrence but, also, such operations are indicative of an earlier failure of deterrence.

Defining ‘non-state actors’ should be less problematic. On first appearances, however, it may seem an unnecessarily broad description for this paper. The category is expansive and deterring through threat of force is hardly an appropriate strategy for interactions with transnational corporations, such as Wal-Mart, or non-governmental organisations such as Médecins Sans Frontières. ‘Terrorism’ may be a more direct and accurate description. But there are problems with the term, not least that it is laden with values that could prove unhelpful when considering deterrence. Nor does ‘terrorism’ adequately cover the breadth of non-state actors such as transnational criminal gangs or 21\(^{st}\) Century anarchists. Consequently this paper will use the term ‘non-state actor’ to engender unconstrained thought on the subject. Emphasis will, however, sit firmly on the part of the spectrum in which organisations use collective violence to achieve their goals.\(^{13}\)

Thinking about deterrence in the 21\(^{st}\) Century, where it has occurred, has exercised a wide range of academic disciplines as well as practitioners of policy. Each thinker has her or his own explicit opinions as well as implicit biases and agendas. Each also sees different purposes to the study: from a security practitioner who wants to know what airport security has to achieve to dissuade an aspiring hijacker; to a strategist who seeks to understand whether it is conceptually possible to deter the ‘undeterrable’? Rather than isolate these dynamics, the approach this paper takes seeks to embrace them: it is futile, indeed counter-productive, to seek a universally accepted perspective. Instead, this study will construct a dialogue between notional individuals representing the viewpoints of an historian, a security practitioner, a sociologist, a statesman, a strategist, a philosopher and a psychologist. The narrative mode will be a pale imitation of the style of Cicero’s rhetoric\(^{14}\), with the *dramatis personae* identified in italics and their opinions phrased in the third person present tense. Views will not always represent the cross-section of a discipline. Indeed, in some instances, they will draw from the margins in order to create worthwhile discussion.

The dialogue aims at hypothesis by elimination. It also addresses the tension between theory and practice. The rigours of a *Philosopher* unpicking a hypothesis, of peer review for the *Strategist*, and of laboratory work for the *Psychologist* have great utility. But the true value of their work is only realised when their concepts, proven or proposed, are translated into actionable strategy; as an American would describe it, ‘operationalized’. As this is rarely done by the theorist, there is a danger that competence on one side of the fence encourages selective attention to the other side. Many of the theories surrounding deterrence are contested, contradictory or highly bespoke and there is a risk of cherry-picking or over-simplification. A dialogue between perspectives aims to acknowledge and exploit such tensions. It will do so in three segments: of what relevance to contemporary challenges is ‘classic deterrence’ thinking; are non-state actors different to states when considering deterrence and, if so, how; and is it possible to deter non-state actors and, if so, how?

This paper will examine the ‘classic deterrence’ of the Cold War and before, to ascertain its utility as a foundation for thinking about the strategy’s applicability in the 21\(^{st}\) Century. It will surmise that non-state actors are rational and can be deterred. A lack of identifiable strategic cultures and a poor understanding of what motivates different non-state actors’ behaviour will challenge the formulation of a deterrence policy. The paper will argue the importance for effective deterrence of accurately communicating the ‘bargain’. It will also propose that deterrence should be aimed at intentions rather than capabilities and, in the long term, at changing the norms in a relationship. To consider the strategy’s future, one must first examine its past. Ned Lebow sets the challenge by describing Cold War deterrence as having “assumed the status of a religious relic which, if frequently paraded in public and blessed in the right language by the high priests of national security, might protect the community against the ultimate evil of nuclear
war". Whether this is an accurate summary and what the implications are for deterring non-state actors will be the starting point for the dialogue of perspectives.

**What relevance classic deterrence?**

In retrospect, deterrence during the Cold War was relatively straightforward. Both Eastern and Western Blocs were invested in the success of the strategy. Failure would not only signal an irrevocable slide towards defeat for one side, but also nuclear destruction on a global scale. It was presumed that decision-makers would act rationally and, in the final analysis, the mutual interest of survival within a nuclear context would trump ideological differences. In contrast to the 21st Century, Cold War adversaries understood each other better. As a former Vice Chief of the US Air Force put it; “the growling, nuclear-armed Soviet bear was relatively easy to understand and deal with. What is not easy to understand and respond to are the many ‘smaller’ threats – and opportunities – that stem from global integration”. When the Berlin Wall collapsed in November 1989 so too did certainties about strategies tailored to suit the specifics of the Cold War. As states grappled with the new world order and the role of non-state actors, the utility of such strategies was questioned. After 9/11, US Vice President Cheney said; “No treaty or arms control agreement or strategy of deterrence will end this conflict. This is a new kind of war, against a new kind of enemy”. For many, the world had fundamentally changed and policy makers of the early 21st Century found little value in strategies to which the previous generation had entrusted their lives.

In this chapter, the Historian, Statesman, Strategist, Philosopher and Psychologist will discuss whether it is right to dismiss ‘classic deterrence’ as anachronistic, or if it still has value when considering 21st Century deterrence. The Statesman and the Strategist will propose that ‘classic deterrence’ is too flawed to be used as a foundation for contemporary thinking. As Ned Lebow warns, policy makers can become; “too wedded to established beliefs and defend images long after they have lost their utility”. The Historian and the Psychologist will present some of the defects of ‘classic deterrence’, but argue that it still has much to offer when considering contemporary deterrence. The Philosopher will defend the classic foundation of the strategy on both its conceptual and historical merits. The chapter will conclude that ‘classic deterrence’ is a useful starting point when considering whether and how non-state actors can be deterred. The orthodox strategy does, however, need to be updated by thought that is not constrained by the temporal, cultural and technological specifics of previous deterrence practices such as those of the Cold War. The dialogue begins with the question of whether the end of the Cold War has signified a change in the nature of warfare.

Globalisation, it has been suggested, has caused a decline in the dominance of the state in the international system which, in turn, has fundamentally altered the nature of warfare. The Strategist proposes that not only is this the case, but that the same dynamic undermines the utility of deterrence. Using Clausewitz’s construct, ‘government’ as the rational element of the Remarkable Trinity, balances the influences of ‘violence’ and ‘chance’. As globalisation has diluted the role of the state, or ‘government’, so contemporary conflict is a much less rational undertaking. This means that deterrence is less viable because potential belligerents will be less motivated by political costs and benefits. Instead of a political purpose, war will be irrational and unstructured, and delivered by actors other than states who will be driven by a combination of religious fanaticism, technology, ethnicity and culture. In ‘new wars’, the Strategist states, the tools of ‘old wars’, such as deterrence, are irrelevant.

The Philosopher contests the Strategist’s premise that Clausewitz’s theories are based on the primacy of the state and on the inherent rationality of ‘government’. More importantly he argues that, although the characteristics of warfare may be changing in the 21st Century, the nature of conflict will not. As Clausewitz observed; “every age [has] its own kind of war, its own limiting conditions, and its own peculiar preconceptions” but these are issues of character. Even though post-Cold War conflict may be fought by forces as disparate as non-territorial terrorists, clan-based irregular forces, transnational criminal organisations and warlords as well as state armies, the Philosopher finds little evidence that such actors are irrational or apolitical. For non-state actors, the objectives sought by using violence are usually
as political as those of their state peers. Conflict in the 21st Century has lost none of its Clausewitzian traits, nor has its nature changed.

While the Statesman is, like the Philosopher, less convinced that the fundamental nature of warfare has altered, what has changed is that competition with non-state adversaries has become intrinsically asymmetric. The asymmetry of ‘ends’ and ‘ways’ is best illustrated by the Irish Republican Army (IRA) statement after the 1984 bombing of the Grand Hotel in Brighton; “today we were unlucky but remember we only have to be lucky once”. Recent history is also littered with examples of non-state actors using asymmetric ‘means’. Because of this asymmetry, proposes the Statesman, deterrence cannot be mutual. Unlike ‘classic deterrence’, the state may choose to deter, as part of its strategy to not lose, but there is no evidence of a non-state challenger employing deterrence. Another departure from the certainties of Cold War policies, is in the different challenges of deterrence’s two constituent parts: capability/credibility; and communication. In the Cold War, greater effort was expended in generating, in the adversary’s eyes, the credibility that one could and would use nuclear force; communicating the bargain was less challenging. When deterring non-state actors the balance is reversed. The lack of Mutually Assured Destruction makes credibility to use force more straightforward: the ‘means’ through which to do so and how to communicate the bargain between actors of such different character is more problematic. For these reasons, the Statesman suggests the tenets of ‘classic deterrence’ should be used with caution.

Be cautious of assumptions that deterrence worked well in the Cold War warns the Historian. Chinese intervention in the Korean War was not deterred by US nuclear capability. The NATO doctrine of ‘flexible response’ assumed escalation, but records show that Warsaw Pact plans relied on the early and unconstrained use of nuclear weapons. Another falsehood from the Cold War, proposes the Historian, is that deterrence is a military-led activity. When considering the strategy in the 21st Century, the maligned Donald Rumsfeld understood this. He said:

_We have entered an era where enemies are in small cells scattered across the globe. Yet America’s forces continue to be arranged essentially to fight large armies, navies, and air forces, and in support of an approach – static deterrence – that does not apply to enemies who have no territories to defend and no treaties to honor._

The Historian offers that the ‘classic deterrence’ of the Cold War is not as sound a foundation as the Philosopher has suggested.

The dialogue is trending towards equating ‘classic deterrence’ with Cold War deterrence. This, corrects the Philosopher, is myopic. The nuclear age undoubtedly caused the threat of what Schelling termed “pure, unconstructive, unacquisitive pain and damage” to give weight to, and accelerate, thinking about deterrence. But the concept predates the nuclear age, as evidenced by the Roman philosophy of _si vis pacem, para bellum_. Deterrence, he argues, is as relevant today against non-state actors as it has been throughout the history of conflict.

Despite its long heritage, the Psychologist suggests that deterrence as a concept may be specious. ‘Psychological reactance’ is a basic response to the instinctive feeling that one is being restricted or manipulated. All threat-based strategies are, therefore, inherently risky and may incite the very action they aim to deter. For example, in 1961 Khrushchev concluded that restraint would be perceived as frailty and would incur further provocation, but a harder line would encourage Kennedy to greater caution. The Psychologist argues that the difficulty with deterrence, ‘classic’ or contemporary, is that the challenger has free choice in how to respond to the bargain. Assumptions can be made about internal drivers of decision-making, appetite for risk, and how those dynamics can be influenced, but the challenger can never be controlled. Deterrence will only work if the challenger chooses to be deterred, a choice which can never be guaranteed. The Historian develops the point in describing the challenges of studying deterrence. A strategy whose success is defined by inaction can only be judged by counter-factual argument; what would have happened if the actor was not deterred? Evidence of failures of
deterrence will be more prevalent than evidence of its successes. But the proposition that deterrence did not work is as unprovable as the proposition that it did. Hence, both the Historian and the Psychologist caution against assumptions that ‘classical deterrence’ is a proven strategy.

The Statesman explores whether any of the practical elements of Cold War deterrence have utility when considering whether and how non-state actors can be deterred. Morgan observed that deterrence between the US and USSR assumed the features of a system in its own right, to the extent that it was often described as a ‘regime’.33 This suggests that when both sides practice deterrence, there is a shared understanding of the difficulties of interdependent decision-making that will still be applicable in the 21st Century. The Statesman is reminded of her comments on asymmetry by the Strategist who argues that there is no interdependence in unilateral deterrence. The Statesman responds that even a one-sided deterrence regime would still be a positive influence on the state’s own behaviour. As Freedman observes; “during the 1960s Secretary of Defense Robert McNamara operated in an explicitly didactic mode in an attempt to educate [the US leadership] … in what he believed to be the optimum behaviour for the nuclear age”.34 A significant lesson from ‘classic deterrence’ is the importance of changing not only the behaviours of the adversary, but also one’s own behaviour in order to conduct strategic relationships without resorting to unnecessary or disproportionate violence.

Accepting, for argument’s sake the Philosopher’s proposition that the character of conflict may have changed but its nature has not, the Strategist is concerned that ‘classic deterrence’ has a poor track record of addressing the complexity of relations with non-state actors. There are three areas in which the Strategist believes deterrence has failed historically, and which a 21st Century review of the concept must consider. The first are political failings, specifically in misunderstanding the limits of deterrence. As Stein and Ned Lebow state; “from Masada to the Irish Easter Rising … history records countless stories of peoples who waged costly struggles with little or no expectation of success”.35 Echoing Thucydides, fear, anger and individual honour have, in the past, proved more persuasive stimuli for action than dispassionate cost/benefit calculations have proved for inaction. Policy makers have mistaken the symptoms of an actor’s grievance for its causes and, therefore, incorrectly chosen deterrence as a strategy. In ‘classic deterrence’, such misdiagnoses of the situation, according to the Strategist, have, at best, had no effect. In the worst cases36, such mistakes have aggravated the challenger’s sense of isolation or weakness resulting in a violent outcome.

The second broad area of historical failure is the practical problems, of which the greatest according to the Strategist is communication. During the Cold War, technical and cultural compatibilities between the adversaries made for considerably more effective communication than would be the case with a non-state actor in the 21st Century. “Theories of deterrence assume that everyone understands … the meaning of barking guard dogs, barbed wire, and ‘No Trespassing’ signs.”37 As the Cuban Missile Crisis example given earlier illustrates, even in the Cold War this understanding has not always existed. It is the context in which the bargain of deterrence is interpreted that gives the signs meaning, more so than the meaning intended by the sender of the deterring communication. The Strategist is concerned that ‘classic deterrence’ does not adequately acknowledge the increased risk for misunderstanding of communication between state and non-state actors. The third area of historical difficulty lies in the psychology of deterrence and, in particular, the assumption in ‘classic deterrence’ of rational choice. Here, the Strategist defers to the Psychologist to expand the detail.

That ‘rational choice theory’ is now considered an imperfect model of human decision-making devalues ‘classic deterrence’ according to the Psychologist. Founded on ‘expected utility theory’ and modified by economic models38, the basic assumption of ‘rational choice theory’ is that choice is predictably influenced by perceptions of costs and benefits. Schelling applied this idea to deterrence theory, concluding that a rational actor would weigh up the potential costs, benefits and risks before making a decision on that calculation. Although this union of ideas advanced deterrence theory considerably, the model’s utility has subsequently become questioned. “Actors in strategic games, which are characterized by interdependent, reciprocal expectations, do not act exclusively in the light of objective circumstances.”39 A central problem is the underlying sociological assumptions of shared (Western)
norms and values. Commenting on the 1973 Yom Kippur War, Kissinger observed that the US notion of rational behaviour failed to explain why Egypt and Syria would start a war with Israel that they could not win, but would do so to restore pride. These flaws do not irrevocably undermine the broader concept of deterrence. But, the Psychologist suggests, not only was ‘rational choice theory’ considered problematic during the Cold War, but its utility will continue to be challenged when considering how to deter non-Western non-state actors.

Although deterrence may have been effective between states in the Cold War, the Strategist summarises that this is no guarantee that it will work against non-state actors. Historical examples of political, practical and psychological failure of deterrence should cause greater circumspection about the concept. As Jervis comments; “people pay more attention to what has happened than to why it happened. Thus learning is superficial, overgeneralized, and based on post hoc ergo propter hoc reasoning”. The Strategist argues that the efficacy of ‘classic deterrence’ is limited to a set of very precise circumstances. The challenger must be motivated by the potential for advantage and not the fear of disadvantage, be able to restrain all elements of the organisation and, most importantly, consider the organisation or cause vulnerable to the threats of the deterrer. In the gamut of non-state actors, this combination of circumstances might be rare and it would invite failure to apply broadly the lessons of ‘classic deterrence’ to the 21st Century.

The Statesman questions whether Cold War methods, or ‘means’, of deterrence are applicable to non-state actors. The threat of limited nuclear war would not be appropriate to dissuade cyber-attack and it is difficult to imagine how it would deter a terrorist organisation from deploying a ‘dirty bomb’. A more fundamental flaw in ‘classic deterrence’ is that, in deterring non-state actors, one is faced with the dilemma that you are damned if you do respond to a challenger, and damned if you don’t. Rather than a cost to be avoided, a reprisal may be exactly what the non-state challenger wants. In Adler’s view, “deterrence … which was the key strategy for maintaining international security during the Cold War, cannot serve the same purpose in the post-Cold War world. In fact, deterrence establishes and sustains the dilemma in the first place.” The Statesman argues that it is strategic complexity which creates this ‘double-damned’ scenario. If the distribution of power and knowledge is broadly symmetrically between a limited number of culturally homogenous and mutually deterring actors, ‘classic deterrence’ has every chance of succeeding, just as it did between the US and USSR. Conversely, the Statesman is concerned that the complexity of deterring non-state actors goes beyond the designs of Cold War practises.

The Philosopher concurs, but makes the case again that there is much to draw from history if lessons are not to be re-learned. The principles of ‘classic deterrence’ still have utility because they transcend new technologies, such as nuclear weapons. It is a concept unfettered by the specifics of previous conflicts because, regardless of the nature of the protagonists, deterrence is about modifying their behaviour. The end of the Cold War did not mark the end of deterrence as a broadly viable strategy, just of a bespoke style of deterrence. The resurgence of non-state actors indicates to the Philosopher that, in future, deterrence will have to be applied in different ways and enacted through different means.

Having unpicked ‘rational choice theory’, the final contribution in this section of the dialogue falls to the Psychologist, who expands on an alternative explanation of decision-making that has updated ‘classic deterrence’, and is a crucial baseline when thinking about the deterrence of non-state actors. ‘Rational choice theory’ failed to account for the fact that decision-makers are emotional human beings rather than mathematically-modelled rational calculators. As Ned Lebow summarises, humans are “beset by doubts and uncertainties, struggle with incongruous longings, antipathies, and loyalties, and are reluctant to make irrevocable choices”. The Psychologist also argues that it is essential, when considering the diversity of non-state actors, to remember that there is not a universal method of processing information or a standard-issue brain. Instead, intellects are individual and methods of reasoning potentially unique dependant on each person’s specific hormones and chemistry. Furthermore, responses to perceived threats or to gratification through reward (both being the foundations of coercion) originate in primitive instincts designed for survival; they may not be subject to rational decision-making. In detailing the relevance of this when considering deterrence, Scheber details
how “these systems bypass the cerebral cortex, which includes portions of the brain responsible for analytic reasoning and self-awareness. In other words, the rational actor may never get a vote when these systems are triggered”. The implication is not that deterrence is impossible, but that a more nuanced understanding of psychological responses is required. To illustrate this argument, the Psychologist offers the briefest case study of the emotional response of fear.

Fear is one of the most important motivating emotions in deterrence, but it can generate irrationality in decision-making. Because fear played an evolutionary role in ensuring individual and species survival, the pathways in the brain that react to fear are better established than the more recently developed faculties of reasoning. In other words, fear produces physical responses before the brain consciously perceives and considers the threat. In offering an example of fear changing societal behaviour, the Psychologist pre-empts criticism that phenomena which may govern individual behaviour on the savannah, or contemporary equivalent, is not relevant in explaining the behaviour of non-state actors in the 21st Century. After 9/11, over 1 million Americans choose to travel by car rather than aircraft. An article by Szalavitz concluded that “driving is far more dangerous than flying, and the decision to switch caused roughly 1,000 additional auto fatalities, according to two separate analyses comparing traffic patterns in late 2001 to those the year before.” Instead of ‘rational choice theory’, a better basis from which to understand such behaviour, the Psychologist suggests, is the concept of ‘bounded rationality’. This idea incorporates observed anomalies and heuristics, and proposes that choices made do not always reflect the highest expected payoff. ‘Prospect theory’ is an economic model of ‘bounded rationality theory’ for decision-making in environments of risk, and it is this model that has the greatest utility for thinking about deterring non-state actors.

The authors of ‘prospect theory’, Tversky and Kahneman, demonstrated in multiple controlled experiments that “most people systematically violate all of the basic axioms of subjective expected utility theory in their actual decision-making behaviour at least some of the time”. Instead, they proposed a theory founded on situation specific analysis. Risk-taking behaviour is based not on the psychology of an individual decision-maker, but on how an interpretation of the situation constrains that individual’s options. Paraphrased by the Psychologist, people frame decisions depending on their perception of their situation, and evaluate alternatives differently depending on the ‘frame’. If, for example, a decision-maker feels that she has lost something (that her current status is undesirable) she exists in what ‘prospect theory’ terms the ‘domain of losses’. The decision-maker will be more accepting of risk as she attempts to regain that which was lost. Conversely, in the ‘domain of gains’, her behaviour will be more risk averse. Although principally an economic model, ‘prospect theory’ has clear relevance for deterrence. As Tversky and Kahneman state; “although the present paper has been concerned mainly with monetary outcomes, the theory is readily applicable to choices involving other attributes … [and] where the probabilities of outcomes are not explicitly given”. The Psychologist proposes that in ‘prospect theory’ lies the means to better understand decision-making in response to a specific communication of deterrence. The theory even obliquely suggests that if one can influence how the challenger frames its options before making a choice about a deterrence bargain, its decision can also be influenced. ‘Prospect theory’ is, therefore, critical when thinking about deterrence and non-state actors.

Much ground has been covered in examining the history of deterrence. This chapter concludes that, with important caveats, ‘classic deterrence’ is a valid starting point to think about the deterrence of non-state actors. For the purposes of this paper deterrence should be seen as a theory, not a doctrine. As Clausewitz reminds us; “whenever an activity deals primarily with the same things, even though there may be minor variations and an infinite diversity of combinations – these things are susceptible to rational study. It is precisely that inquiry which is the most essential part of any theory”. What is required is not a rejection, but a revision of ‘classic’ deterrence theory to make it applicable to the changed characters of adversary and conflict.
Are non-state actors different?

It would appear axiomatic that non-state actors are different from state actors; if they were the same, they would not require a different classification. But there are significant implications for deterrence in this bold assertion. Deterrence theory makes an underpinning assumption that both actors share a common view of what the costs and benefits of the bargain are. Is this assumption sound when one of the actors is transnational, without the apparatus of statehood and motivated by different drivers than its state adversary? As Adler posits; “deterrence only works when the rules of the game and norms of appropriate behavior are intersubjectively understood and shared”.52 It is entirely plausible that non-state actors, such as terrorist groups, and the radical states that may sponsor them, lack that common understanding of deterrence. They may also reject the norms of the established international relations behaviour. In attempting to deter non-state actors, lack of common language in strategy is not the only challenge. Unless there is a profound understanding of the non-state adversary, there is a risk that the bargain of deterrence will be inelegantly tailored or ineffectively communicated. For these reasons it is essential to explore whether non-state actors are fundamentally different from states and, if so, how.

In this chapter, the Sociologist, Statesman, Strategist, Philosopher and Psychologist will explore those differences between state and non-state actors which have a bearing on the strategy of deterrence. Generalisations will be necessary because of the multitude and diversity of non-state actors. The deterrence iconoclast Ned Lebow observed that “threat-based strategies involve an equation. On one side is the ability to inflict punishment; on the other is the ability to absorb costs. Theorists have emphasized the former and all but ignored the latter”.53 In order to address this shortfall of attention to the non-state challenger, the Philosopher will first question the assumption that they are fundamentally different. The Psychologist and Sociologist will offer ways in which state and non-state actors do differ, before the discussion will broaden to address the question of whether the latter are rational. The Strategist will raise the issue of strategic culture in non-state adversaries. The discussion will finish with an exploration of state sponsorship and specific motivators of non-state actor behaviour.

By the end of the chapter, the dialogue’s participants will determine that although non-state actors are not a new phenomenon, current understanding of them might be inadequate for effective deterrence. Another conclusion will be that all non-state actors can be deterred because they are all rational decision-makers. Notions of rationality, however, may differ markedly from Western concepts. The strategic cultures of non-state actors will also differ from their state counterparts, but such cultures do exist even if they are hard to identify. The final deduction will be that the challenges created by the fundamental difference between state and non-state actors include communicating the deterring threat, operating in relation to the norms and conventions of the international system, and the use of global media.

It is a flawed assumption, according to the Philosopher, that violent non-state actors are a new or a very different type of adversary. As Jervis states; “it is not surprising that terrorism is ancient, because individuals have never been fully bound to states and terror is needed by the weak, who lack other instruments”.54 While 21st Century globalisation may have created a new generation of non-state actors, most that threaten state security, such as transnational organised crime or ideologically motivated regional insurgencies, have a longer pedigree. The Philosopher suggests that they are better understood than might first be thought and strategies concerning them have historical precedence on which to build in the 21st Century. The Psychologist agrees that this is not a new epoch of adversary, but cautions that understanding the specific nature of the non-state adversary is critical for effective deterrence. On both sides of the deterrence equation, the interpretation of word and deed depends on frames of reference. These are “laden with cultural and psychological overtones and passed through overlapping prisms of history, culture, language, ideological axioms, and modes of transmission and reception of information on the ‘other’.”55 The Sociologist offers that understanding, particularly of terrorism, is inadequate. For example, social sciences have traditionally given little regard to religion (possibly because the majority of social scientists are not religious56), yet this is fundamental to understanding the motivations of many non-state actors.
Religion, the *Sociologist* suggests, can affect a non-state actor’s susceptibility to deterrence in two ways. Indirectly, it can change attitudes towards risk. Religions that credit their high-priests with unquestionable authority, or terrorist leaders whose religious beliefs and that of their followers impute the same infallibility, will be markedly resistant to deterrence. Such religions can also create a different perspective of the value of human casualties. A more direct method by which, according to the *Sociologist*, religion impacts on deterrence is that it provides a version of reality which may differ from that of the deterring state. Belief in a higher order and spiritual intervention informs a calculation of costs and benefits in which “the reward for obedience to divine will and the punishment for disobedience – occurring in the hereafter – transcends any earthly punishment that the enemy can inflict”. Such perspectives serve to undermine the neat Western logic of offer of threat and bargain on which state-on-state deterrence has relied. Religion is one example of the aspects of non-state actors that state designers of deterrence policies have not, historically, understood in sufficient depth.

Another possible difference between state and non-state actors proposed by the *Psychologist* is the degree of choice each has, an essential ingredient for deterrence. There is a presumption that, for example, a terrorist group can be dissuaded from using terror, “but it is often precisely the absence of choice (embodied in the belief that political grievances cannot be redressed through conventional means) which forces terrorists to adopt the methods they do”. This, the *Statesman* responds, might lead to a conclusion that terrorists are ‘undeterrable’, but that is too broad a statement. There are certainly radical perspectives derived from religion (for example, al-Qaeda) or apocalyptic beliefs (such as Aum Shinrikyo). These extreme views are manifested in seeking unattainable objectives outside the norms of international relations and may, according to the *Statesman*, be impervious to deterrence. But there are different types of terrorist groups; organisations such as the IRA or Basque Fatherland and Liberty (ETA) constrain their actions to achieve limited political goals without undermining sympathy for their respective cause. It is vital to conceptually separate those who are open to incentives and, therefore, can be deterred, from the likes of Aum Shinrikyo who cannot. The *Strategist* concurs, arguing that the actions of many non-state actors “can no longer be described as rational political activity. This is because the rational elements of the Clausewitzian Trinity, the military and principally the government, are no longer present”. Whether the rationality of non-state actors is different from that of states is central to the discussion of whether the former can be deterred.

The *Philosopher* strongly disagrees with the proposition that irrational acts are indicative of irrational decision-making which, in turn, makes non-state actors ‘undeterrable’. Clausewitz’s concept of war as a rational instrument to promote political interests did not, in his view, preclude irrational behaviour in conflict. Handel commented that; “as Clausewitz emphasises, war is in all aspects a reflection of human nature, rife with non-rational motives”. Thinkers, such as Creveld, propose that the 21st Century marks a shift, or return, to stateless warfare. In response the *Philosopher* suggests that although brutality and plundering were features of medieval conflict, they were conducted with strategic goals in mind and do not indicate a lack of political rationality. The *Psychologist* agrees that non-state actors are just as rational (or irrational) as any other actor. A more plausible explanation is that behaviour may intentionally appear irrational so as to discombobulate state adversaries, or that it reflects a construct of rationality that differs from the Western ideas of costs and benefits. The noteworthy success of terrorist networks and transnational criminal organisations in the 21st Century is borne of their rational behaviour in exploiting the opportunities, and resources, created by globalisation. The decision-making of such actors can “be understood in classic Clausewitzian terms … terrorism is a highly instrumental activity”. Both *Philosopher* and *Psychologist* agree that all non-state actors are rational decision-makers, but that their concepts of rationality may be not be as familiar as that of state paradigms.

Accepting that proposition, the dialogue moves on to discuss strategic culture as the embodiment of a non-state actor’s specific type of rationality. The problem, suggests the *Strategist*, is that non-state actors have no appreciable strategic culture in the Cold War sense. Yet, in that conflict, it was an understanding of such cultures that proved the essential ingredient to successful deterrence. What is necessary in the
21st Century, he suggests, is greater emphasis on symbols and myths, and greater comprehension of the global significance of these phenomena as they are communicated through modern media. With this expanded understanding, the concept of strategic culture will be more easily applied to non-state actors. The challenge for the practitioners of contemporary deterrence, highlights the Strategist, is that it took decades for the Cold War adversaries to truly understand each other. Conversely, in the 21st Century there are a host of non-state actors who might need to be deterred, each with a different strategic culture; a more complex and time-consuming proposition. A difference between the generic strategic culture of terrorist actors and states, suggests the Psychologist, may exist in the tolerance for organisational pain. Motivated by strong political, social, or cultural stimuli, terrorist organisations also often have long-standing grievances, a combination of which makes for a powerful driving force. This is problematic for deterrence as the threat of “even the most powerful punitive measures may have no deterrent effect if the terrorists feel no pain”. In essence, actors, such as terrorists, have little to lose and all to gain.

In addition to the challenges of a lack of common framework for strategic culture, another variance between state and non-state actors, submits the Psychologist, is the mechanisms and norms of communication. Communicating the threat and bargain of deterrence is fraught with opportunities for misperceptions or mistakes, as illustrated by the phenomenon of ‘biased escalation’. An experiment required two people to apply mechanical pressure to each other’s fingers with instructions to reciprocate the pressure received. In reality, what happened was that the pressure applied was escalated. The reason for this is that anticipated force when pushing down on one’s own finger seems less than the unanticipated forces when another pushes down on one’s finger. This experiment is evidence, the Psychologist argues, that self-produced sensations are less salient than those produced by others. Social groups, be they state or non-state actors, will be more aware of pain suffered at the hands of ‘the other’ than the pain they inflict on ‘the other’. When considering deterrence, there is a risk, therefore, of under-estimating the impact of the threat when articulating the bargain.

The Statesman is not so convinced that state and non-state actors are immutably different. Both exist in an international system that is still defined by statehood and, more pertinently, many non-state actors are reliant on state sponsors to mitigate the weaknesses of statelessness. This reality affects well-defined and state-based groups, such as Hamas, as well as more diffused actors, such as al-Qaeda. As Jervis suggests; “September 11 and its sequelæ show the crucial role of states … al-Qaeda gained many of its capabilities through its capture of the Afghan government. It could not have operated as it did without the acquiescence of the state”. States that sponsor non-state groups can be subjected to traditional state-on-state deterrence using established and well-rehearsed methods. Such activities will deter the non-state actor by proxy, a proposition far less fraught with practical challenges than attempting to deter then directly. In summary, the Statesman concludes that the reliance of some transnational groups on the physical aspects of statehood is evidence that non-state actors are not wholly different from their state adversaries.

Another area in which non-state actors differ from states, suggests the Strategist, is that the former are not bound by the norms and conventions of the international system. On the contrary, non-state actors often use those norms to constrain their state adversaries. Reflecting their own construct of rationality, transnational terrorist organisations have “consistently used and abused the constitutive norms of the international order for instrumental reasons, while at the same time condemning their victims for not abiding by them”. In a similar vein, the Statesman argues that some non-state challengers actively seek a violent reaction from a state adversary. Such a strategic aim would be borne of a desire to increase the importance of the non-state actor, to provide a post hoc mandate for its action or simply to promote group unity. These dynamics might make deterrence, and certainly compulsion, a counter-productive strategy.

The final difference between state and non-state actors is one proposed by the Sociologist and concerns their use of the media. The globalisation and digitisation of media enables non-state actors to circumvent many limitations of their lack of state structure. These changes also make a deterring state’s
actions visible and accountable to an extent, and in ways not previously experienced. As Gray observed, a media with global reach and real-time access “means a political and cultural-moral audit of behaviour that will be an enduring feature of future strategic history”.74 Gaining publicity through the global media is often one of the non-state actor’s objectives. Even the tactical failure of a terrorist attack, such as in Mumbai, therefore, can be defined as a strategic success if it generates electronic ‘column inches’. Common to both state and non-state actors is that compelling activity in front of a global audience determines, to a greater extent, that audience’s view of whether a party is ‘winning’ or ‘losing’. This perception, argues the Sociologist, is significant in deciding the reality of defeat or victory. The implication of this for deterrence is that the impact of the ‘bargain’ may be diluted by the less controllable dynamic of global opinion.

In conclusion, the dialogue has reached consensus that non-state actors are different and need to be thought about differently. They are not, however, radically dissimilar to state actors and is it important to not label them irrational or ‘undeterrable’. Those seeking to deter non-state actors need a better understanding of non-Western motivations and bespoke strategic cultures, challenging though it might be to acquire this knowledge. How the communication of a deterring threat is perceived, the constraints of international norms and conventions, and the role and exploitation of global 21st Century media are all key differences between state and non-state actors which have impact in the realm of deterrence. Ultimately, as with all actors, those without statehood consist of human decision-makers so the basic psychology of deterrence, and its concomitant pitfalls, continue to apply.

How does one deter non-state actors?

Violence committed by non-state actors for political reasons exists on a spectrum that ranges from civil disorder to acts of international terrorism, such as 9/11. Police presence at a rally might deter the former; how to deter the latter is more challenging. The protester will exercise a degree of self-restraint and broadly respect the principles of democracy or the rule of law. Through the excesses of violence and choice of victims, a terrorist shows no respect for either.75 Despite the challenges of finding solutions, having accepted in this paper that non-state actors can be deterred, how to do so must be explored. The greatest value of a theoretical discourse lies in its practical applications. Tangible outputs of a policy of deterrence, such as the medium by which the bargain is offered and to whom, will therefore be examined. The previous chapter examined the ways in which the character of non-state actors differed from their state adversaries. The premise of this chapter is more focussed on questions akin to ‘what is the al-Qaeda equivalent of the World Trade Centre?’76

Little detail can be found in official literature about the specifics of delivering deterrence. Writing just before the publication of the UK’s 2010 National Security Strategy, Strachan made the following observation.

“Current British defence policy says very little about its deterrent functions – little, that is, by comparison with any statement on defence issued between 1957 and 1991, and with what is still said by the United States in its quadrennial defence reviews”.77

There are obvious security reasons for this. It might also be explained as a reflection of the success of deterrence. It could also be indicative of the complexity of delivering such a strategy. A study of the history of deterrence also fails to offer clues as to which tactics could succeed and which might fail, even for the less discreet strategy of compulsion. For example, some have suggested that, despite a spike in ‘revenge activity’ in the immediate aftermath, the US airstrikes on Tripoli in 1986 resulted in Libya becoming more constrained in its sponsorship of international terrorism.78 Although written in 1996, that hypothesis does not appear to consider in its evidence the 1988 bombing of Pan Am Flight 103 over Lockerbie. There is more evidence that the US airstrikes failed to moderate Libyan sponsorship to any significant degree.79 Despite an inconclusive historical catalogue and limited contemporary doctrinal guidance on the subject, this paper will still attempt to identify some of the methods by which non-state actors can be deterred.
In this chapter, all previous participants in the dialogue will be joined by a Security Practitioner, and will discuss the effectiveness of various practical options to deter non-state actors. The Security Practitioner will offer for debate two examples of effective deterrence of non-state actors. The Strategist introduces the idea of deterring the intentions rather than the capabilities of non-state adversaries. Achieving broad agreement with this proposal, the Philosopher will make a subtle but critical refinement; that deterrence’s ultimate goal should be to change the norms, in addition to the behaviour of non-state actors. The real-world utility of ‘prospect theory’ will be contested by the Strategist and the Psychologist. After discussing the strengths and weaknesses of various other methods of deterrence, the dialogue will concentrate on a detailed exploration of the importance of communication. The chapter will conclude that it is difficult to identify commonly applicable tactics as non-state actors can be so different in character. Understanding these differences and then tailoring the tactics of a deterrence strategy is vital. The success or failure to do so will be made manifest in the ‘bargain’ and how it is communicated; the indication of success being that the bargain is perceived by the non-state challenger as it is intended to be by the state deterress. The most important deduction is that the execution of 21st Century deterrence against non-state actors should seek a different end-state to that of its Cold War progenitor. Deterrence of non-state actors must be aimed at changing the norms of the relationship between the two parties so that, eventually, deterrence itself is no longer necessary.

Even with the most extreme example of non-state actor violence – the suicide bomber – the Security Practitioner argues that deterrence is possible. An example from the Occupied Territories in 2003 is used to illustrate this proposal. An Israeli agent telephoned an affluent Palestinian businessman who lived in the Gaza Strip to inform him that his son had, for several months, been making preparations for a suicide bomb mission. The agent told the businessman that if his son went ahead, he and his family would be held liable for the consequences. Specifically their house would be knocked down, Israel would sever all commercial links to the merchant’s business and he and his family would be barred from ever entering Israel. The businessman confronted his son and successfully persuaded him that the costs to his immediate kin of becoming a suicide bomber outweighed any benefits to the Palestinian cause. A good understanding of the correct mix of threats, argues the Security Practitioner, not only enables deterrence of individual ‘foot-soldiers’ but also the decision-makers within non-state actors. Since its withdrawal from Lebanon in 2000, Israel has employed a mixture of compulsion and deterrence to constrain attacks by Hizballah. Although a qualified success in many areas, the 2006 war did mean that “Hizballah is not eager to go another round with the IDF [Israeli Defence Force]. Hizballah is a quasi-government in that it controls territory and provides services to local residents. Thus its leadership feels pressure when its people suffer.” The Security Practitioner offers Hizballah’s silence, when Israeli forces entered Gaza in 2008, as evidence of successful deterrence.

The Psychologist questions the long term effectiveness of such tactics. They may dissuade temporarily or make attacks on Israel from the Occupied Territories more difficult to mount, but have they persuaded organisations such as Hizballah and Hamas to give up the strategy of an armed struggle? The Historian suggests that what the Security Practitioner describes is not relevant to the debate. The degree of success of the wider Israeli strategy against its state and non-state adversaries is not just deterrence but also a large degree of compulsion. To describe the use of military force to alter behaviour that has already started as deterrence is inaccurate. Moreover, it is the strategy of compulsion that has had greater effect against Israel’s state adversaries. As the originator of the Security Practitioner’s example goes on to admit, although the Israeli approach of blending deterrence and compulsion has “shown extraordinarily good results in deterring states from launching direct attacks against Israel, it is too early to judge its level of success against guerrilla and terrorist operations.” The Psychologist and Historian argue that the increase in the use of violence by Palestinian groups, reaching a crescendo in the Second Intifada, is evidence that Israel’s strategy has failed to comprehensively deter its non-state adversaries.

Switching tack, the Security Practitioner offers examples of how the practical manifestations of deterrence had an effect on the behaviour of non-state actors. The introduction of ‘sky marshals’ in 1971 and 100% baggage checks at airports in 1973 reduced aircraft hijackings from a peak of 91 in 1969, to just 19 in
1986. After successful ‘hostage rescue’ operations in the 1980s, such as that demonstrated by the UK Special Air Service in the Iranian Embassy siege, non-state actors who might have chosen to advance their political aims through these specific tactics were deterred from doing so. The Strategist contests this trend analysis by stating that such evidence is too focussed on targets. The Security Practitioner’s observations measure success in terms of reduction in the frequency of specific terrorist tactics against certain types of targets as a result of bespoke countermeasures. Failing to assess the totality of terrorist incidents obscures the fact that instead of being deterred from using violence by such countermeasures, non-state actors change either their targets or their methods. Evaluating the 9/11 attacks through this lens, the Strategist concludes that not only did ‘sky marshals’ and 100% baggage checks fail to deter those involved, but also that the tactic was an evolutionary combination of the traditional techniques of aircraft hijacking and suicide bombing. The evidence presented by the Security Practitioner of effective deterrence, the Strategist argues, is actually evidence of the adaptability of violent non-state actors.

The Strategist argues that instead of being target-focussed, deterring a non-state actor’s intentions will have a more profound and longer lasting effect. In practise this requires an undermining by the state of the adversary’s belief in its ability to generate a desired effect. Doing so avoids the trap of heightening the adversary’s frustration by successfully deterring its ‘means’ of attack. This frustration is counter-productive, as it manifests itself in renewed resolve and more imaginative efforts to overcome the obstacles of a target-focussed strategy. The Historian finds the Strategist’s argument more compelling than the Security Practitioner’s examples, but offers caveats. While a non-state actor may ‘design around’ a specific countermeasure, this should be seen as positive progress, albeit incremental. As an historical example, Kennan’s design for ‘containment’ is offered. “Deterring Soviet expansion, [Kennan argued], would lead to either Soviet shifts toward normal behavior or collapse of the regime; it could not go on indefinitely pursuing plans impossible to carry out given the West’s opposition”. The Historian also cautions that, on occasion, there may be a need to deter an attack on a specific target or against a specific ‘means’. This will require target-focussed rather than intention-focussed tactics. In summary, while the long term goals of deterrence are more important, the need for short term focussed deterrence should not be dismissed.

In the nuances of the Strategist’s proposal, the Statesman finds value. It is possible, the latter suggests, to further divide the non-state actor’s intentions into two elements: the desire to use violence; and the expectation that the specific act of violence will successfully deliver a political aim. There may be benefit in tailoring deterrence policies to the distinctly different motivations of those who just want to kill and those who want to kill to achieve a strategic outcome. The Philosopher disagrees: it is very difficult to externally influence a person’s or an organisation’s raisons d’être. Trying to deter a violent non-state actor from using violence is “roughly equivalent to demanding that terrorists stop pursuing their larger goals … the question then, is how nations might deter incidents of terror while simultaneously acknowledging that deterrence itself cannot address the reasons why terrorists act in the first place.” Deterring intentions rather than (or as well as) capabilities is a laudable concept, but there is little indication that it works in reality. Israel has long sought to end terrorist activity using all of the elements of coercion, yet continues to be attacked by terrorists because they continue to believe in their objective of Israel’s demise. No strategy can change such a deeply valued strategic objective. What deterrence might be able to do, suggests the Philosopher, is change the norms of the relationship between state and non-state actor.

In the introduction to this paper it was proposed that all deterrence is self-deterrence because an actor has to choose to be deterred. This premise forms the foundation of the Philosopher’s argument for a 21st Century adaptation of deterrence, to focus on norms as well as behaviours. What are required for non-state actors, he argues, are strategies that create internal and habitual deterrence rather than policies that address specific intentions. As Freedman phrases it; “a norms-based approach requires reinforcing certain values to the point where it is well understood that they must not be violated”. To change norms, the perception of potential costs incurred by pursuing a specific course of action needs to become wider and more entrenched. Over time this will preclude a wider range of options to the non-state adversary than just the one that the original policy aimed to deter. Relatively straightforward to
explain, the Philosopher concedes, but less so to deliver. The signalling of strong normative disapproval for a non-state actor’s intentions or capabilities is proposed by some, but this may be more suited to the captive and dependent audience encountered in raising children than to international relations. Another way of affecting norms is to change the social construct in which the relationship between state deterrer and non-state challenger exists. For example, identities, interests and practices common to both sides can be altered. In essence, these actions are designed to alter the ‘rules of the game’ in a bilateral relationship before violence occurs.

The Psychologist interjects that there is a clear and very important link in this line of argument to ‘prospect theory’. The frame for decision-making that divides the domains of gain and loss is defined as much by the social constructs of the relationship as it is by whether the decision-maker ‘feels’ she is winning or losing. Changing norms is an activity that does not sit wholly within the realm of deterrence, but equally a deterrence strategy that does not have this goal in mind might preclude any advances in this area. While limited to state actors, the example of Gorbachev’s policies of glasnost and perestroika, and the effect they had on the Cold War, illustrates this theory. The Philosopher accepts that the success of this constructivist proposal will rely on a very specific set of circumstances. He also accepts that the full range of coercive techniques will probably be required to create such a set of conditions, as evidenced by Kennan’s containment policy and its successors. The key point, though, is that changing norms should be the ultimate goal of any deterrence policy. The Statesman approves this point and suggests that, while the Philosopher’s explanation has not produced any practical outputs which prescribe how to deter non-state actors, it has clearly indicated how not to coerce them. For example, the 1988 US cruise missile strike against a Sudanese pharmaceutical factory (which was allegedly producing pre-cursors for VX gas for Iraq) was also intended to compel the Sudanese government to stop its support of, or punish it for historical ties with, al-Qaeda. This blunt use of force not only punished Sudan, despite its already changed behaviour (it expelled bin Laden in 1996), but reinforced a negative relationship between the actors instead of developing more constructive norms.

Changing the balance of the cost/benefit analysis is not only a vital component of successful deterrence, proposes the Strategist, but also a useful step towards changing the norms of behaviour. A non-state actor may perceive that the advantage of using violence to achieve political objectives is that it incurs less risk because its lack of statehood makes it less vulnerable. The first stage to change the ratio of cost and benefit for the non-state actor is to make it more difficult to have an effect by increasing defensive measures, in essence to ‘harden’ possible targets. The next step is to deny the non-state actor its goal of justifying its mandate through an equally violent and indiscriminate response to the provocation. Retaliation, if it must happen, should avoid allowing “terrorists to increase their performative power in front of local and global audiences and delegitimize their adversaries”. Humiliating an opponent can only be counter-productive in the long term. Conversely, argues the Strategist, acknowledging the challenger’s dignity could increase its sensitivity towards benefits and, therefore, change the cost/benefit balance in the deterrer’s favour.

Understanding in advance a non-state actor’s propensity for risk is another area of difficulty. The Statesman suggests that, while ‘prospect theory’s’ notion of a ‘frame’ is very useful to understand conceptually how perceptions of gain or loss affect decision-making, it is challenging to assess where in real life that ‘frame’ may sit. Judgement on attitudes towards risk can be made using a relative comparison to other choices available after the deterrence ‘bargain’ has been communicated. It is extremely difficult to do so in advance, when the policy for a specific non-state actor is being formulated. The Statesman questions whether an analyst could actually “apply any kind of psychological theory to a complex environment where significant variables cannot be carefully controlled or measured as they would be in a laboratory setting.” This does not devalue the role of such theories in helping to shed light on decision-making in the general sense. In producing tangible outputs for a specific deterrence policy, however, the activity of framing is susceptible to ‘judgemental heuristics’ which can have a significant effect on how options are presented to the challenger. Moreover, the biases and heuristics of individual decision-makers in non-state organisations mean that ‘prospect theory’, or other similar models, do not offer a ready and easily applied solution. The
Psychologist, surprisingly, appears to agree. Human brains are not identical; genetic changes throughout generations of evolution have created diversity in behaviour which policy makers should attempt to understand. To make this task even harder, disease, physical injury, chemical or hormonal imbalances, as well as external phenomena such as ‘group-think’, contrive to make tactics based on consistent and predictable human behaviour questionable.

However, ‘prospect theory’ does not pretend to answer all the questions about decision-making in all circumstances, reminds the Psychologist. It does recognise that individual character traits will have an impact. Most importantly, the model reinforces the importance of the context in which decisions are made. Deciding which domain an actor perceives itself to be in can be made objectively or subjectively. What is interesting for the discussion about deterring non-state actors, suggests the Psychologist, is whether they will predominately perceive themselves to be in the domain of losses because they exist outside of the established construct of the international system. Prospect theory interpreted in an international relations context suggests that leaders take more risks to regain their international positions, reputations, and domestic support than they do to maintain or enhance the same. Consequently, adversary leaders are more readily deterred from taking an action than they are compelled to stop or reverse an action. Acknowledging that “deterrence is an inherently uncertain business,” the Psychologist proposes that ‘prospect theory’ should be the starting point for understanding the context in which different non-state challengers will react to a deterring state.

Picking up on the inherent uncertainty of deterrence, the Security Practitioner argues that it can, therefore, be only part of the answer about how to influence non-state actors. While the long term aspiration for deterrence may be to create conformity to the norms of peaceful international relations, it has a better (though still debatable) track record of short term dissuasion of specific violent attacks. But even in those cases, deterrence must never be relied upon as the single solution. Instead, it should be regarded as one element of multi-faceted strategy and “because irrational or unpredictable behaviour is likely to be a factor, there must be a capable and effective response option.” The Historian agrees. Case studies of deterrence, such as the United States’ efforts to deal with Barbary pirates or the Russian conflict with Chechen nationalists, show that punitive and deterring strategies often overlap in practice. Indeed, in these examples, this multi-layered approach was necessary to address the specific character and motivations of the respective non-state adversaries.

The Strategist offers another practical suggestion for policies concerning violent non-state actors; that the deterring state needs to consider its definition of success. Even when a terrorist incident occurs, if its scope, duration or intensity is constrained because of deterrence then the strategy should be considered a partial triumph. This is important, not only to devalue the adversary’s definition of success, but also to have an assessment of the effectiveness of particular tactics of deterrence. This presumes, according to the Philosopher, that one could ascertain the full extent of the terrorist organisation’s desired outcome and, therefore, the scope of partial failure. He does agree, though, that society’s frame of reference should be addressed, although he cannot suggest how. Violent activities by non-state organisations have an exaggerated effect because society is less habituated to war, health epidemics or widespread civil disorder. Using 9/11 as an example, the Security Practitioner supports the Philosopher by stating that; “modern societies may … be uniquely vulnerable psychologically. The density of personal networks multiplies the number of people who lost a close friend or who know someone who did. The rapid flow of information also means that everyone immediately learns about any terrorist attack, and because everyone gets this information, it dominates not only the thoughts of separated individuals but … is incorporated into popular consciousness and culture.”

Returning to the discussion about the ‘means’ by which non-state actors can be deterred, the Statesman asks how one knows when one has successfully deterred? In addition to the difficulties of knowing what aspect of one’s policy affected which decision-maker’s behaviour, there are the challenges of timescale. An absence of terrorist activity may be temporary. Only when there is a level and duration of commitment to non-violent discourse, for example as shown by the IRA, can deterrence, or whatever coercive strategy was employed, be declared a success.
Not only is it difficult to know when deterrence has been successful but, the Security Practitioner offers, it is also hard to know which capability has deterred a non-state actor. Unless the threat presented to a non-state actor as a consequence of choosing action has been very specific, it is difficult to identify the “usefulness of systems whose utility rests, paradoxically, on their non-use”. In an effort to broaden the discussion beyond military capabilities, the Historian presents the equivocal case for economic threats as part of the deterrence ‘bargain’. The threat of sanctions may provide some motivation for a non-state actor to change its behaviour, but few have economic structures to which sanctions could be applied. The tactic may be more applicable to state sponsors of violent non-state organisations, yet there is the received wisdom that sanctions cause a population to suffer more quickly and more profoundly than they cause decision-makers to change behaviour. The Historian suggests that, whilst it is still too early to be definitive, current events in the Ukraine may prove instructive. The compelling economic pressure applied to Russian power brokers in order to affect the behaviour of the non-state Russian separatists may change perceptions on the nature and utility of economic levers. An opposing view is that affecting economic resilience might obliquely strengthen a non-state actor. For example, “a semi-criminal government like that of Slobodan Milošević, with a good range of underworld connections, thrives in an environment of smuggling and illegality”. A final area of risk with economic deterrence is that, if sanctions are subsequently imposed, they may perpetuate the grievances and discord in the challenger’s societal base and, as a consequence, popular support for the non-state actor would be hardened or increased.

The rationality of non-state actors has already been discussed, but the Statesman comments that a concept of rationality is when an actor’s political and psychological realities converge in the perceptions of others. When they differ, and it is the psychological reality that defines behaviour, that behaviour is seen as irrational. A practical extrapolation of this view is that, within a non-state organisation, there will be the fanatics willing to sacrifice their life or freedom for ‘the cause’. Their psychological reality differs from their political reality and this makes the communication of deterrence a challenge, if not an impossibility. Above these fanatical ‘foot soldiers’ sit the power brokers who are likely to be more rational in that their political and psychological realities are perceived to be aligned. It is the power brokers or decision-makers against whom deterrence should be focussed. For the Statesman, the most critical component in deterring non-state actors lies in effective communication. It is not just a case of getting the message right, when all of the differences of strategic culture conspire against that, but also of correctly identifying to whom the message should be sent.

The Philosopher agrees; “even in societies that nurture support for suicide bombers, there are rational actors who can exert influence over these seemingly undeterrable individuals.” Conversely, one should not underestimate the challenges of finding common purpose with a violent non-state adversary, cautions the Security Practitioner. Evidence is offered of an Irish diplomat who, in the 1980s, was in involved in the ‘back-door’ dialogue between the British government and Republican elements in Northern Ireland.

“The effect of that on the Provisionals [IRA] was, of course, to encourage them to become even more intensive in their campaign of violence. In other words, every time this [dialogue] happened, going back to the earliest days of The Troubles, the Provisional IRA took encouragement from the fact that the British were reaching out to them”.

The lesson for deterrence, according to the Security Practitioner, is that it is not just a question of communicating the right message to the right people, but doing so at the right time.

To deter a non-state actor, agrees the Psychologist, one must understand its frame of reference. Those constructing a bespoke deterrence policy need to consider what options are available to the adversary’s decision-makers. The social and ideological perspectives through which they will interpret the deterrent message must be appreciated. Perhaps most important is to accurately assess what the decision-makers’ individual, and the organisation’s collective, motivations are. Do they seek increased power, financial
gain or global recognition? Or, more prosaically, do they desire security or are they fighting for survival? “Those seeking to implement a successful counterterrorist deterrence strategy . . . need to understand the world as it appears to the terrorist.” The Historian argues that the fact that many of al-Qaeda’s core leaders have been killed suggests that threats to life do not deter them; denial of martyrdom may be a more deterring proposition with which to bargain. The Psychologist offers that it is not just an understanding of the frame of reference of the non-state actor that is important, but also how the bargain is framed. To illustrate the point, the laboratory example is presented of two options for a hypothetical flu vaccination programme to deal with an epidemic expected to kill 600. The first group is told that Policy A will save 200 lives. The alternative, Policy B, has a \( \frac{1}{3} \) chance that 600 lives will be saved, but a \( \frac{2}{3} \) chance that no one will be saved. 72% chose Policy A. A second group is told Policy A will cause 400 to die, but Policy B has a \( \frac{1}{3} \) chance that nobody will die and a \( \frac{2}{3} \) chance that 600 will die. 78% chose policy B. In the subtleties of communication between culturally different actors, states the Psychologist, correctly framing the ‘bargain’ is vital and could define the success or failure of deterrence.

The Sociologist agrees wholeheartedly with the centrality of communication when considering how to deter non-state actors, but cautions that what is easy to describe may be much harder to deliver. The pitfalls of perceived challenges to honour have already been detailed. Many non-state actors place little value in the bureaucratic arrangements found in the state construct they have rejected, or that has rejected them. Instead they place more value in kinship in various forms. Carefully crafted threats to kinship groups might be a useful ‘bargain’ for deterrence, but inadvertent threats could be disastrous. The importance of how the deterring message is phrased, to whom it is given and when, have been identified. Also important, argues the Sociologist, is who imparts the message, the personal relationship between messenger and non-state decision-maker and the form of the message. All of these recommendations are plausible and sensible, but the impact of real world challenges must not be lost. To illustrate the point, the Sociologist recalls a comment by Sir Bernard Ingham, Margaret Thatcher’s Chief Press Secretary, about dealing with the IRA:

Interviewer: “Do you think either of you actually understood the Irish [Republican perspective] any more in 1990 than you did in 1979?”

Ingham: “Probably not. No, I don’t make any claims about that. But they sure didn’t understand us.”

This sobering example is also useful, as it shifts focus away from the decision-making elite back to the role in deterrence of the wider organisation. In the case of non-state actors like al-Qaeda, absolute faith of the ‘foot soldiers’ in their creed or cause will not only make them harder to deter individually, it may also constrain the decision-makers whom they serve, even if those decision-makers are less fundamentally motivated. Such a dynamic could, however, also present opportunities for deterrence. The Sociologist suggests that religion, for example, can be “exploited to enhance deterrence. Superstitions can induce courage and self-sacrifice on the personal level, but may also induce fear and mass hysteria on the collective level”. The complexity of exploiting a dynamic like this, and of getting the message right, makes communication the most difficult, but also the most important aspect of deterring violent non-state actors.

Those that present the more hypothetical aspects of this dialogue may be disappointed that the theory of deterrence appears not to offer an exhaustive panacea of practical outputs. On the contrary, the Historian, Security Practitioner and Statesman know that real life is messy, and actual deterrence policies not only have to be tailored to each recipient, but are inevitably compromised by reality. There is no unequivocal and universal set of deterrence tactics that could be used against violent non-state actors. As Payne advises; “deterrence is an art and uncertainties in its functioning are unavoidable in the best of circumstances.” One of the greatest challenges is guarding against biases. Democracies tend towards a presumption of the universality of democratic values, that all differences can be negotiated. Analysts and policy makers must guard against the influence of their own biases and ideological perspectives when developing deterrence policies. This chapter has, however, reached agreement in some areas
about how to deter non-state actors. Deterring intentions rather than capabilities is likely to be more successful in the medium term. In the long term, deterrence policies should aim at changing the norms of the relationship between the deterrer and the non-state challenger. The most critical aspect of deterring non-state actors is communication of the bargain; the correct message, given to the right decision-maker, at the right time, by the right messenger and framed in the right way, is key to successful deterrence.

Closing thoughts

Prescription is neither the stated aim of this paper nor a realistic proposition for a topic with as ill-defined boundaries as the deterrence of non-state actors. Rather than concluding remarks, therefore, a summary will be presented of the ground covered by this dialogue of perspectives. First, it will be useful to revise the context of the debate. States can exploit three main sources of influence: diplomatic, economic and military (or a combination thereof). As Cooper neatly summarises; “they can persuade, they can bribe or they can coerce. At first sight persuasion seems the weakest of the three. In so far as words carry weight, it is because they represent either promises of assistance or threats of violence.” However, both economic controls and military force have a blemished track record, particularly when applied to the trickier proposition of non-state actors. These tricky stateless entities that pursue political goals through the use of violence are not unique to the 21st Century. Contemporary globalisation has, however, diminished the disadvantages they suffer due to their lack of levers of influence in comparison to states. Asymmetry is the result and has caused a re-evaluation of well-established strategies. Despite deterrence’s pedigree – that it “actually has been ‘done’, certainly attempted, since strategic history began” – there is little precedence for how to apply it to non-state actors.

Clausewitz argued that an aversion to shedding an enemy’s blood is an exploitable weakness, but it is from the work of Sun Tzu that deterrence takes philosophical sustenance. “To subdue the enemy without fighting is the acme of skill.” While the concept of deterrence can be condensed into a Twitter post, the evolution of its practices has been less straightforward, particularly regarding non-state actors. In the Cold War, the rigid laws of ‘rational choice theory’ were superseded. While ‘rationalists’ and psychologists seek to explain the same thing, how they do so differs. As Mercer observes, ‘rationalists’ use “deduction, statistics, and probability theory, whereas psychologists rely on induction and description of how the mind actually works”. ‘Prospect theory’ incorporates subjective values into the ‘classic deterrence’ equation, but a refined understanding of the psychology of deterrence does not necessarily produce a clearer notion of what capabilities deter. As Gray comments; “we refer to the deterrent, an assertion of unilateral merit over what is surely a relational variable”. Only the non-state adversary can decide what it is that deters, and it is only the adversary who can choose to be deterred. The psychology is, however, important. Having considered the challenges of deterrence perhaps more than most military leaders, Admiral Mullen reminds those grappling with the issue today that; “more than 40 years ago, Henry Kissinger warned that deterrence is ‘above all a psychological problem. The assessment of risks on which it depends becomes less and less precise in the face of weapons of unprecedented novelty and destructiveness’”. The wisdom of Kissinger’s view is unaltered by the 21st Century resurgence in threats posed by non-state actors.

Much of the dialogue in this paper has trended to the theoretical; absolutes are more easily articulated in the abstract than in the practical policies. This tension between theory and practice has exercised many minds, not least that of Plato. In his Simile of the Cave, he concluded that “society will never be properly governed either by the uneducated, who have no knowledge of the truth, or by those who are allowed to spend all their lives in purely intellectual pursuits”. In explicitly contrasting perspectives of ‘thinkers’ and ‘practitioners’, this paper has aspired to cause the former to descend into the unenlightened reality of ‘the cave’ and view deterrence through what Clausewitz described as the light of reason refracted by factors outside of the academic domain. Even if such grounded enlightenment is not achieved, there is still value in the process of dialogue. Cooper offers the challenge that, “whereas professional soldiers try, almost as a matter of routine, to learn from the past, there is little equivalent effort in the diplomatic
world". In thinking about the deterrence of non-state actors, the least that can be gained is the reflected insight into the deterrer’s own decision-making assumptions, ideology and strategic culture.

Before summarising the main thrusts of the dialogue, it would be remiss not to thank the Historian, Security Practitioner, Sociologist and Statesman, the Strategist, Philosopher and the Psychologist for sharing their thoughts and engaging with the debate. In their wide-ranging discussion they reached tentative consensus in the following areas. While acknowledging its flaws, ‘classic deterrence’ is a useful foundation when considering how to apply the strategy to non-state actors. Crucially, those actors are rational in their use of violence for political purposes and thus they can be deterred. What makes deterrence significantly more challenging in the 21st Century than it was in the Cold War is the quantity of – and variety in – adversaries’ strategic cultures. This is not just a conceptual challenge; a profound understanding of the culture and motivations of the non-state challenger is critical if the threat and bargain of deterrence is to be effectively communicated. The dialogue concluded that target-focused deterrence may have a role in the short term, but there is greater value in policies aimed at deterring intentions rather than capabilities. In deterring non-state actors, the long term goal must be to change norms. Deterrence’s greatest achievement would be to take an actor, alienated from the traditions of the international system and with no inherent customs of constructive behaviour, and change those norms such that any form of coercion would be unnecessary in the future.

Endnotes

12 Adler, “Damned If You Do”, 211.
14 Chapter 5 of Michael von Albrecht, Cicero’s Style: A Synopsis (Boston MA: Brill, 2003) gives a comprehensive study.
28 Schelling, Arms and Influence, 16.
33 Morgan, “Taking the Long View”, 751.
34 Freedman, “Deterrence: A Reply”, 793.
35 Ned Lebow and Stein, “Beyond Deterrence”, 125.
36 For example, US policies regarding Japan which contributed to the latter’s decision to attack Pearl Harbour in 1941. A fuller exposition is in Ned Lebow and Stein, “Beyond Deterrence”, 128.
37 ibid.,133.
39 Adler, “Damned If You Do”, 203.
43 Ned Lebow, Between Peace and War, 107.
50 Kahneman and Tversky, “Prospect Theory”, 288.
51 Clausewitz, On War, 162-163.
56 Jervis, American Foreign Policy, 37.
57 Bar, “God, Nations and Deterrence”, 449.
58 ibid., 448.
60 Bowen, “Deterring Mass-Casualty Terrorism”, 27.
61 Fleming, “New or Old Wars?”, 218.
63 Fleming, “New or Old Wars?”, 225.
64 Adler, “Damned If You Do”, 211.
67 Ilardi, “Deterring Terrorism”, 43.
71 Jervis, American Foreign Policy, 41.
72 Adler, “Damned If You Do”, 208.
76 Bowen, “Deterring Mass-Casualty Terrorism”, 27.
82 Cooper, Breaking of Nations, 121.
85 Ilardi, “Deterring Terrorism”, 44.
88 Ilardi, “Deterring Terrorism”, 41.
89 Carter, “Transcending the Nuclear Framework”, 89.
90 Freedman, Deterrence, 4.
94 Adler, “Damned If You Do”, 221-222.
95 McDermott, Risk Taking, 180.
96 Ibid., 182 – 183.
97 Ibid., 6.
98 Scheber, “Evolutionary Psychology”, 462.
99 Ibid., 469.
100 McDermott, Risk Taking, 35.
101 Payne, “Understanding Deterrence”, 403.
104 Jervis, American Foreign Policy, 39.
106 Cooper, Breaking of Nations, 118.
107 Ibid.
111 Ilardi, “Deterring Terrorism”, 41.
113 McDermott, Risk Taking, 21.
119 Cooper, Breaking of Nations, 116.
125 Mullen, “It’s Time for a New Deterrence Model”, 3.
127 Clausewitz, On War, 133.
128 Cooper, Deterrence, 85.
Complexity, war, and the complexity of war

Lt Col D J Bickers MBE RE

CHAPTER 1 – INTRODUCTION

In his foreword to *British Defence Doctrine*, the Chief of Defence Staff states that warfare is “an increasingly complex venture”. In many ways military doctrine, and the academic thinking on which it draws, has elevated complexity to one of the defining qualities of the contemporary and future operating environments. Complexity, or derivatives of it, is mentioned in *British Defence Doctrine* as more than two of the three fundamental elements of war identified by Clausewitz. Its prominence continues throughout subordinate doctrine, with its recognition being listed as a guiding principle of campaigning. *Global Strategic Trends*, the Ministry of Defence’s comprehensive view of the future out to 2040, presents an apparently inexorable march towards increasing complexity. There is almost no trend, from climate change to global inequality, from population growth to resource scarcity, which does not add complexity to the world and, in turn, to warfare. Conversely, not a single identified trend makes the world less complex.

But does arguing that future warfare is increasingly complex risk implying that past warfare was not? Writing on war would suggest that complexity has been a constant feature. Michael Handel argues that there was a profound increase in the complexity of warfare in the mid-Nineteenth Century due to the technological developments of that age. An inability to see through this complexity, he says, is a factor in no strategic thinker emerging on a par with Clausewitz or Sun Tzu after that time. And yet Clausewitz and Sun Tzu also considered warfare to be uncalculatedly complex in their own times. Clausewitz described how efforts to deduce a universal set of principles or rules for war were frustrated by the “endless complexities involved”, which he blamed on the inherently interactive nature of war. Before Clausewitz, Sun Tzu also observed that the process of interaction brought incalculable complexity, saying that the interaction of two forces meant that “their combinations are limitless; none can comprehend them all.”

So, if dynamic interaction has always made war complex, what use is it to speak of increasing complexity? Could complexity be an absolute quality whose threshold was breached in war many centuries ago? And, therefore, is talk of increasing complexity essentially meaningless? Michael Gallagher and others have argued that the assumption of an unprecedented level of contemporary complexity is both “descriptively false” and of limited analytical usefulness. Clearly what is needed, if the increasing complexity of warfare is to be proved and if complexity as a descriptor is to have any meaning beyond a mere synonym for the complicated and unpredictable, is a measure of complexity, a means to quantify its prevalence within war. But how do we measure complexity?

Measurement can only follow definition. Later we draw on some of the detailed properties of complexity identified by complexity theory. Before then it is sufficient to say that complexity is a quality of a network or system and, to draw on a simple definition, such as that of John Miller and Scott Page, who describe a complex system as “a set of diverse actors who dynamically interact with one another awash in a sea of feedbacks.” Complexity thus becomes a function of actors and their interactions. We would, therefore, expect a measure of complexity to include measures of these two attributes. Vincent Vesterby concurs with this, stating that, “Complexity is quantity and diversity of components and relations. Anything that increases these factors increases complexity.” However, he also concludes that complexity has qualitative aspects which do not render themselves amenable to pure quantification. This difficulty with agreeing a universal measure of complexity is acknowledged throughout the literature on complexity. Stewart Umphrey, for instance, questions biologists claims of an “evolution of complexity” because of their inability to quantify levels of biotic complexity. So here we hit our first hurdle in verifying the increased complexity of warfare, namely that there is no agreed way to quantify it.
It is this conundrum, an assertion that warfare is becoming increasingly complex but a seeming inability to quantify that increase, which forms the foundation of this paper. The paper attempts to give meaning to this increased complexity and draw implications from it. As complexity resists pure quantification, the paper begins by identifying the qualitative factors which underpin claims of increasing complexity. This is done in Chapter Two where a review of literature on contemporary war identifies the factors, namely globalisation and the information age, which have been argued to increase war’s complexity. Chapter Two also introduces complexity theory, which is used as a theoretical framework throughout the remainder of the paper. Chapter Three combines the findings of the literature review on contemporary war with complexity theory. It seeks to give meaning to, and draw implications from, the increased complexity of war at the strategic level. Whereas this discussion in Chapter Three is necessarily abstract, Chapter Four draws more concrete conclusions for campaigning at the operational level. Chapter Five summarises and concludes the paper.

The main thesis of the paper is that war increased in complexity during the 1990s due to the effects of globalisation and the information age. This increased complexity can be understood by an increase in the number and diversity of actors who influence warfare, and the extent to which they are interconnected. The result is that the organisational topology of war has changed. Clausewitz’s dyadic conception of war is an increasingly inaccurate model. His analogy of war as a duel, typified by his assertion that “a picture of it as a whole can be formed by imagining a pair of wrestlers” fails to capture the diversity of strategic actors who now influence contemporary warfare. A better model is to conceptualise war as a multi-actor complex network. The paper challenges the suitability of both mainstream international relations and risk management theories for dealing with the universal dynamics underpinning complex phenomenon. It also highlights why the military, as a tool of national strategy, will be increasingly perceived as both imprecise and indecisive. The paper contends that understanding warfare as a complex network can help explain our ability (or indeed inability) to exert control at the operational level. It argues that understanding operations as a competition for network control is the most useful way to plan and execute campaigns. It suggests an ability to observe and understand network dynamics will be key to exerting any degree of control, and that the data revolution will have significant implications for how this is achieved.

Methodology

The paper’s ontological position is most closely aligned to that of critical realism. Critical realism is a post-positivist ontology concerned with the discovery of the unobservable currents that guide, but do not determine, historical events. In this sense it closely matches the aims of complexity theory. It can accommodate and seek explanations for unobservable phenomenon (such as emergent system behaviours, a defining characteristic of complexity) in a way that positivism cannot. This paper uses multiple research strategies. Initially, a literature review frames the phenomenon of complexity in war and identifies the main themes emerging from it. These themes are then analysed through the lens of complexity theory. The investigation is predominantly qualitative, but is facilitated by some quantitative analysis of data on complexity in war.

CHAPTER 2 – LITERATURE REVIEW

This chapter reviews the literature on contemporary war to identify reasons attributed for its increased complexity. It also outlines the historiography and foundational concepts of complexity theory so we can employ it in later chapters to analyse more deeply the increasing complexity of war.

The increased complexity of contemporary war: the ‘new war’ thesis

We have established that complexity defies simple quantification. How then has contemporary warfare been judged as qualitatively more complex than that which went before? The past shows that the evolving character of war is driven most acutely by factors emerging not from war itself, but from the
wider societies and economies from which it emerges. MacGregor Knox and Williamson Murray, in charting the dynamics of military affairs, identify five military revolutions. The first three are all driven by profound societal change (the creation of the nation state, the Industrial Revolution and the French Revolution) and the fourth (the First World War) was driven by the legacies of the second and third. Only the fifth, the advent of nuclear weapons, can claim not to have its roots in deep-seated societal change. So it would seem likely that any increase in the complexity of warfare would be caused by an increase in the complexity of wider society. This is certainly the view of Christopher Coker, who states that “by the early twentieth century war was becoming more dangerous because it was becoming more complex; and its increasing complexity mirrored the increasing complexity of life.”

So to account for the increased complexity of warfare we need to account for the increased complexity of life. Here there is one overriding explanation: globalisation. The definition, timelines, spread and implications of globalisation are all contested. A widely held view is that contemporary globalisation (for many argue there have been earlier manifestations) was sparked by the invention of the microchip in the 1970s and fuelled by the subsequent proliferations of computers and the internet. A simple definition of globalisation is given by Ian Clark; “movements in both the intensity and the extent of international interactions.” So we can see immediately that globalisation impacts directly on one of complexity’s defining attributes, interaction. This increase in interaction can be seen most starkly in the increased level of connectivity that the world has experienced since the 1990s. Another cited impact of globalisation is that it broadened the range of actors who influence international affairs and caused a concomitant reduction in the relative power of nation states. So we can see that, by increasing both the number of international actors and the extent of international interactions, globalisation has qualitatively increased both attributes of complexity identified in Chapter One.

What effect has globalisation had on warfare? One of the earliest writers to identify its implications was Alvin Toffler. His 1980 book, Future Shock, gave a periodized account of history which predicted the arrival of an “information age”. His follow-up book, War and Anti-War, contended that the information age was evoking a similar revolution in the military as it was in business, namely placing a higher premium on information and less on mass. By the 1990s several more writers were arguing that globalisation, in conjunction with the profound structural change in the international system brought about by the end of the Cold War, had fundamentally altered war. Martin Van Creveld argued that war was entering a new “non-Clausewitzian” epoch characterised by different actors (guerrilla armies, terrorists and bandits instead of nation states), methods (irregular instead of regular warfare) and goals (justice and religion instead of a Clausewitzian continuation of policy). Similarly, Mary Kaldor gave globalisation as the context for a new “post-modern” form of warfare characterised by differing goals (identity politics instead of geopolitics), methods (guerrilla warfare over conventional) and economy (decentralised instead of centralised funding). Moving into the next decade, Rupert Smith argued that 1991 had seen a paradigm shift from interstate industrial war to “war amongst the people” and Colin Gray stated that, by reducing state autonomy, globalisation had increased the need for multinational security responses. Andrew Hoskin and Ben O’Loughlin also spoke of a paradigmatic shift in modern warfare. They claimed a “perpetual connectivity” created a new media-ecology which saw the media weaponized and made a tool of warfare. Most recently, Emile Simpson argued that globalisation had replaced war’s “polarity” (the extent to which war is considered a Clausewitzian duel between two actors) with a kaleidoscope of fragmented political dynamics and, in so doing, had made strategy increasingly sensitive to global third-party audiences.

Collectively, these writers can be thought of as advancing a ‘new war’ thesis. Other authors have disputed this thesis, but the debate has largely centred on the extent to which the changes purported by the new war authors affect the nature of war, rather than merely its character. In fact, some of the authors who dispute the new war thesis on these grounds accept its observations as long as the changes it causes to war are only argued to exist in war’s character. For instance, Fleming states that “Though there is a danger that the ‘new war’ idea will become exaggerated, it nevertheless highlights trends that if correct will impact, if not on the nature of war, then certainly on its character and conduct”. So some epistemological semantics apart, the changes identified by the new war thesis have been largely accepted within mainstream western academic thinking.
John Kiszely concludes that these new wars present contemporary military commanders with unprecedented challenges: an increased number of lines of operation to coordinate (diplomatic, economic and social in addition to military), and an increased number of actors to contend with (other-government departments, non-governmental organisations, contractors, and host-nation representatives). The net effect, he claims, has greatly increased the complexity of military command in the post-Cold War era.\(^{31}\) In this regard the changes brought about in war by globalisation mirror those brought about in the international system. In arguing an increase both in the number of actors, who influence war, and the extent of interaction between them, the new war thesis is arguing an increase in both attributes of complexity.

A common view amongst authors is that this increased complexity reduces the utility of war. David Kilcullen describes how increased connectivity results in malign actors becoming increasingly interconnected and nested within licit systems, such that it becomes “virtually impossible to target a dark network without also harming the community within which it nests.”\(^{32}\) Simpson contends that war becomes less decisive as the expansion in the number of actors engaged in warfare dilutes the “notion of definitive victory”, meaning that, at best, we aspire to ‘success’ rather than ‘victory’.\(^{33}\) Similarly, Coker observes that increased complexity presents the possibility that we are headed “inexorably to the inconclusiveness of war”\(^{34}\) which will offer “ever-diminishing returns”.\(^{35}\) James Rosenau even suggests that the “blunting of conventional military approaches” will diminish the extent to which states resort to violence.\(^{36}\) Finally, as pure military responses become less effective, several authors suggest future responses will need to integrate more levers of national power. For instance, Kaldor describes the need for a “cosmopolitan response” which integrates more closely politics, economics and security issues.\(^{37}\)

Does empirical evidence exist to support the new war thesis that war has increased in complexity? One of the only attempts to measure the complexity of war was conducted by Hemda Ben-Yehuda in 2008.\(^{38}\) His intent was to chart complexity over time but compare war’s complexity to the level of accommodation in its eventual outcome. But, by taking the complexity scores he attributed to the 55 wars which took place between 1946 and 2002, and by mapping them over time (Figure 1), we can see whether there is evidence for an increase in complexity. Ben-Yehuda’s measure of complexity was taken as a function of a war’s “issue complexity” (the substance and scope of issues being fought over) and its “structural complexity” (the number and diversity of actors).\(^{39}\)

His results, particularly when viewed by decadal average complexity score (Figure 2), suggest an increase in complexity from the 1990s onwards. But, even ignoring the small sample size, do his results really demonstrate an increase in complexity, particularly as we have already noted the difficulty with quantifying complexity? What Ben-Yehuda is measuring, in essence, are the factors which the new war thesis claims underpin the increase in war’s complexity. The factors which the new war authors use to
justify this claim all attract higher scores under Ben-Yehuda’s methodology. So his findings support their thesis, namely that there was a change in war around the 1990s, but do not, by themselves, advance the argument that war has become more complex. To empirically show the presence of factors claimed by some to increase complexity merely supports the claim that those factors are present, not that they actually increase complexity. To argue that they do would be tautological. He reached no decisive conclusions linking a war’s complexity to the degree of accommodation in its outcome and, therefore, his work neither supports nor refutes the argument that war has reduced in utility.

Future war as increasing complex?

What are the arguments for war continuing to increase in complexity? The most compelling relates to further increases in global connectivity, which in turn will increase complexity. Eric Schmidt, the Chairman of Google, claims that while half the world has access to some combination of cell phone and internet technology, large groups of countries remain either “partially-connected” or in the process of “connecting”. The speed with which such technologies have previously spread would suggest that proliferation to these currently unconnected regions will be rapid. For instance, cell phone usage in Pakistan increased from 300,000 in 2000 to 100 million in 2010. This increase in connectivity is likely to affect not only life but also war, with Kilcullen citing “increasing connectedness” as one of four megatrends which will shape future conflict.

If we are to explore more fully the implications of increasing complexity for war we will need to establish a deeper theoretical understanding of complexity. To do this we turn to complexity theory.

Complexity theory: a very short history

During the 1990s a ‘new science’ of complexity began to emerge, largely as increased computational power was brought to bear on nonlinear mathematics. This new science, which we will group under the umbrella term complexity theory, has the potential to aid understanding of complexity as a phenomenon. Complexity theory is a set of related theories which have emerged from a range of disciplines, which complicates charting its development. However, its major conceptual building blocks are best considered as: general systems theory, cybernetics, chaos theory, complex adaptive systems theory and network theory.

The point of departure for general systems theory is generally taken as the summary work of Ludwig von Bertalanffy, translated into English in 1968. Bertalanffy gave root to complexity theory’s inherently interdisciplinary approach, believing that similarities between disciplines would allow a general theoretical framework to emerge. Amongst Bertalanffy’s insights were the tendency for systems to seek equilibrium, the nonlinear nature of many interactions, and the differences between open and closed systems. Around the same time Cybernetics (a word adopted by Norbert Wiener from the Greek
think of kubernetes, meaning pilot) was also taking a systems approach and stressed the role of information feedbacks loops.47 Chaos theory emerged in the 1980s and described how small changes in initial conditions can have a disproportionate effect on outcomes within a system. It was popularised by the metaphor of the ‘butterfly effect’ and the best-selling book, Chaos, by James Gleick.48 Complex adaptive systems theory investigates how structures and patterns of interaction self-organise from disorder.49 It constitutes an active research field across a range of diverse disciplines, such as statistical physics, physical chemistry, biophysics, geophysics, astrophysics, theoretical ecology, semiconductor physics and optics.50 More recently network theory, spurred by the rapid expansion of the internet, has built on traditional network and graph theory in seeking to understand the behaviour of networks with very large numbers of nodes.51 It is in this field that complexity theory has seen the greatest advances in the last decade, to the extent that some claim it has fundamentally altered the understanding of complexity.52 There are no universally agreed principles for complexity theory, but the following concepts feature most prominently in the literature.

### Complexity theory: foundational concepts

A complex system comprises actors (or elements, agents or nodes) who interact. Feedback is a circular series of interactions which can be positive (in which case the system moves away from equilibrium) or negative (in which case it stabilises).53 A system can be said to be complex rather than merely complicated when the extent of these interactions cause the system to display emergent behaviour, described as properties which the system possesses which its separate agents do not. Robert Axelrod and Michael Cohen use consciousness as an example: no single neuron possesses it, but the human brain does.54 A system which does not display emergent properties, no matter how many actors and interactions it possesses, is merely complicated, not complex. A significant consequence of emergent behaviour is that a complex system is non-reducible, it cannot be understood by analysing its component actors in isolation. As Jervis explains, a complex system is not only greater than the sum of its parts, but different than the sum of its parts.55 Another important feature of a complex system is the nonlinear nature of its interactions, defined by Donella Meadows as “a relationship between two elements in a system where the cause does not produce a proportional (straight-line) effect.”56 Emergence and nonlinearity cause complexity to be synonymous with uncertainty.57 Complex systems are therefore difficult to understand, with the effect of any intervention being exceptionally difficult to predict.

During the 1990s, physicist Per Bak introduced the concept of self-organised criticality, an attempt at a universal explanation of complex phenomenon.58 Self-organised criticality identifies the tendency for complex systems to evolve into a poised but critical state, far from equilibrium and where minor shocks can lead to events of all sizes. This stands in contrast to the earlier work of Bertalanffy who assessed that open systems would “tend towards a steady state”.59 The example Bak used to illustrate this concept (often re-quoted) is of a sand-pile. When sand is dropped to form a pile it will initially form a shallow structure whose behaviour can be understood by the individual properties of the sand grains. Eventually, as the pile becomes steeper and, when the pile is far out of balance, avalanches occur. These avalanches can only be understood by a holistic understanding of the whole pile, not by a reductionist understanding of the individual grains. The sand pile is therefore a complex system. The magnitude of the sand pile’s avalanches will vary from small (and insignificant) to large (and catastrophic). Bak contends that the magnitude of its avalanches are fractal and that, when plotted on a graph, they conform to simple power-law distributions, with the probability of an avalanche’s occurrence decreasing as its magnitude increases. Bak contends that change in many complex systems, from earthquakes, to biological evolution to stock-market fluctuations, conforms to such power-laws. Power-law distributions are described as fat-tailed, meaning that low-magnitude but high-probability events are statistically significant and account for the majority of change. The evolution of a complex system is therefore often perceived as a series of “punctuated equilibriums”.61

Considerable substance has been added to self-organised criticality through network theory’s study of network topology, aided by the availability of large datasets on real complex networks such as the internet. An important finding in this field was made by Albert-László Barabási and Reka Albert who
identified that network expansion occurs through preferential linking, that is new nodes are more likely to attach to a network through an already well-connected node. This was contrary to traditional network theory which assumed random network generation. Whereas randomly generated graphs generate a single giant component (Figure 3a), complex networks form structures with some highly connected nodes (hubs – see Figure 3b) but many more poorly connected nodes. Preferential linking results in complex networks organising into a scale-free state, where the network’s degree distribution (the frequency distribution of the connectedness of nodes within the network) conforms to a power-law, an observation validated through empirical study of complex networks as diverse as proteins and the internet. This diversity hints at universal laws which underpin all complex networks, with scale-free structures argued as the reason why complex networks self-organise to a critical state (the phenomenon identified by Bak).

Scale-free networks display other unique characteristics. First, they have been shown to have a much lower diameter (the average shortest path between any pair of nodes) than equivalent random graphs. They are, in effect, surprisingly highly connected with even very large networks having very small diameters. Second, their structure makes them very resilient to random-attack (nodes being removed at random) but, unsurprisingly, highly vulnerable to targeted attacks on their hubs. Finally, as a complex network’s sub-systems become more tightly interconnected, so change in the network is more likely to be discontinuous or abrupt, due to a series of cascading failures running through these sub-systems.

Complexity theory: a critique

Complexity theory is not without its criticism. Some take the view that complexity is a ‘pop-science’. The fact that complexity theory has captured the popular imagination and resulted in successful book sales for Gleick and Bak does not, of itself, constitute effective criticism. What is more, the strong foundation of the science in a range of respected peer-reviewed scientific journals disputes this line of argument.

A more compelling critique of complexity science is put forward by Giorgio Israel who suggests it has failed to demarcate between ontological and epistemological levels. He argues that the degree of chaos or complexity in a system belongs to the epistemological level of prediction, not to the ontological level. He claims,

“It would be an abuse to consider the existence of chaos in a model – which is a purely mathematical fact – as proof of the chaotic nature of the phenomenon this model is claimed to represent. In this way we would surreptitiously be introducing the idea that the model is a perfect image of the phenomenon being examined.”
This is a valid criticism, and should be used to ensure we keep Alfred Korzybski’s famous premise in mind. Korzybski remarked that “a map is not the territory it represents but, if correct, it has a similar structure to the territory, which accounts for its usefulness”\textsuperscript{74}, meaning that we should not confuse the abstractions through which we observe the world with the world itself. But if the aim of complexity theory is taken as improving the usefulness of our map, then it would seem a perfectly valid endeavour.

The final criticism of complexity theory is that it cannot provide a level of understanding beyond the broad but vague level of metaphor. Writing in 1997, George Johnson suggested that it was unclear whether the concept of complexity would be “something like energy, which can be precisely metered, or something like, say, prudence or charity, which will remain forever in the realm of the subjective.”\textsuperscript{75} In 2005 Israel was still making the same observation, “we are still only half way along the path leading from the vague commonsense notion to a notion defined in rigorous scientific terms but still non-existent.”\textsuperscript{76} However, the view of network theorists is that the unprecedented availability of data offers the chance to map the networks behind complex systems and move complexity theory beyond this vague and subjective position. Such an approach Barabási, the most prominent of the complex network theorists, claims is “fundamentally reshaping our approach to complexity.”\textsuperscript{77}

So complexity theory can serve two purposes. The earlier work, typified by Bak, can help explain complex phenomenon at a metaphorical level. The more recent work, typified by Barabási, offers the hope of understanding complexity at a more objective level. But this work is less mature (much of it being published in the last 12 months) and therefore any deductions drawn from its application to warfare should be considered as tentative until a broader academic base has had time to review its claims.

CHAPTER 3 – COMPLEXITY AND THE STRATEGIC LEVEL OF WAR

This chapter combines the findings of the literature review on contemporary war with complexity theory. It assesses the implications of the increased complexity of contemporary conflict for the strategic level of war.

The applicability of complexity theory to war

Several authors have drawn on complexity theory in the study of international relations. James Rosenau employs it to understand the increased complexity of world affairs, specifically what he considers to be their increasing turbulence.\textsuperscript{78} Robert Jervis draws on emergent behaviour and nonlinear interaction to argue that mainstream political science theories are ineffective at explaining an increasingly interconnected and unpredictable world.\textsuperscript{79} Emilian Kavalski has even suggested that complexity theory, particularly Bak’s concept of self-organised criticality, is so relevant in explaining international relations that it has prompted a “fifth debate” in that field.\textsuperscript{80} While the ferocity of that debate should not be overstated, there is nonetheless a broad body of opinion supporting complexity theory’s usefulness in understanding international relations, and an associated view that it challenges the validity of more conventional meta-theories.

Other authors have drawn on complexity theory to help explain war. Colin Gray strikes a cautious line, writing that “complexity, chaos and nonlinearity are important and valid ideas that are applicable to, indeed are inherent in the structure and dynamics of, strategy and war, but their relevance should not be overstated”.\textsuperscript{81} But others have applied it more fully. Jim Storr argues it gives useful insights into combat\textsuperscript{82} and Damian Popolo similarly that complexity theory, “has the potential to enhance our understanding of social reality in general and of conflict in particular”.\textsuperscript{83} Justin Kelly writing with David Kilcullen\textsuperscript{84} and Zoltán Jobbagy writing with László Szego,\textsuperscript{85} have all stated that war should be considered as a complex adaptive system. So, as with international relations, there is a body of opinion that complexity theory can aid our understanding of war.
War as an enduring product of complexity

Support to the argument that war can be understood through complexity theory is that war displays self-organised criticality, one of its foundational concepts. The frequency and magnitude of war displays fractal properties, a fact first identified in 1948 by Lewis Richardson, well before Bak identified self-organised criticality. Richardson identified that the magnitude and frequency of wars between 1820 and 1945 complied with a power-law distribution giving the “the suspicion that some fairly general tendency concerning aggregation for aggression is revealed by these otherwise scattered phenomena.” In 2003 Lars-Erik Cederman updated Richardson’s work with data covering wars from 1820 to 1997. Cederman also observed a power-law distribution and attributed the phenomenon to self-organised criticality. In Figure 4, Richardson and Cedermans’ analysis has been updated once again by applying data for wars up to 2007.

Figure 4 – Cumulative frequency distribution of severity of inter and intrastate wars, 1818-2007.

Figure 4, in effect, gives a representation of the propensity for war to manifest and escalate, with the horizontal axis serving as a spectrum of conflict, with the violence of war increasing from left to right. The gradient of the graph, which Bak refers to as the fractal dimension, and Mandelbrot as the scaling exponent, is a function of the ratio of large wars to smaller wars and can, therefore, be considered an expression of escalation. Clausewitz constructed a concept of absolute war, an abstract form of war which, totally unrestrained by policy, he said would inevitably escalate through interaction to the most violent confrontation possible. The reason escalation to absolute war did not occur in real war, he argued, was due to the restraint of policy. What is not clear is why escalation in wars across history should comply with a power-law.

Bak presents fluctuations and catastrophes in complex systems as unavoidable laws of nature. But this presents a rather fatalistic outlook, one where the occurrence of war across all ranges of magnitude becomes almost guaranteed as a result of it being a complex phenomenon. Cederman’s explanation for war’s self-organised criticality is based around periods of technological change driving larger wars. But this would seem to contradict Clausewitz who tells us that war’s escalation towards his abstract absolute war is regulated by policy, not technology. Steven Pinker suggests it is the underlying process through which forces of attrition and escalation combine: coalitions can get bigger, wars can last longer and losses can get heavier. This is consistent with Clausewitz’s conception that war in theory would escalate towards an absolute level of violence, but it still does not explain why that relationship should be fractal. Perhaps more convincing is Claudio Cioffi-Revilla and Manus Midlarsky’s argument that it is a result of the international system self-organising to a critical state. If the explanation for earthquakes complying with power-laws is found in the underlying complex interaction of the system from which they originate (the tectonic system) so it would seem most likely that the explanation for war complying with power-laws is found in the underlying complex interaction of the international system. Complexity theory tells
us that if the international system is a complex system then fluctuations in it are inevitable. But it is only fluctuations which are inevitable, not their translation into violence. The extent to which violence is directly proportional to upheaval in the international system depends on the success of policy in stopping it from occurring or, once initiated, from escalating. So war should be considered as a product of complexity within the international system and, while it is not an inevitable product, Figure 4 shows it has nonetheless complied with some remarkably consistent patterns of manifestation throughout history.

Is confirmation that human agency can overcome the self-organising inevitability of war the deviation from the linear trend-line around magnitude six events (one million deaths)? Could it be that some form of agency has prevented as many large-scale wars as the power-law would predict? Nate Silver has applied the same power-law methodology to the frequency and magnitude of terrorist attacks in Israel and notes a similar deviation for high magnitude events. He asserts this is an indication that Israel has succeeded in preventing very large-scale terrorist attacks.\(^96\) Similarly, perhaps the collective strategic choices made by the international system since 1818 have been successful in preventing the number of magnitude six wars that would otherwise have been expected. While it is impossible to prove, this insight does at least support the assertion that complexity need not inevitably lead to violence.

### The complexity of new wars

So having shown that war is a product of external complexity, how can complexity theory also add meaning to war as an internally complex phenomenon? Specifically, how can it help understand the complexity of the new wars identified in Chapter Two, where the number and diversity of actors involved, and the increase in their interactions, have all increased? Furthermore, can it support the claim that these changes have caused war to lose utility and, therefore, require a more integrated political-military response? As discussed in Chapter Two, the most recent advances in complexity theory have been made by network theory, which has sought to explain complex phenomenon through studying the complex networks which underpin them. Thus, to understand the changes brought about by the new wars it would seem logical to take a network perspective.

By taking such a network perspective we can describe the new wars as having affected the topology of war. They have increased both the number and diversity of actors, and their extent of interaction. In so doing they have expanded the network which underpins war, and increased the density of connections within that network. The nub of a war may be a confrontation between just two actors, but the new war thesis would argue that third-party actors exert an influence on the outcome of that war. So let us consider war as a complex network involving many interconnected actors wider than just its two principle belligerents. Network theory can provide insights into the structure and, in turn, dynamics of such a complex system.

No single actor exerts global control over war’s complex network, so its structure can be best understood as resulting from self-organisation. Network theory tells us that complex systems self-organise to a scale-free state. This has several implications. First, as previously noted, scale-free networks are surprisingly highly connected. While the actors in war’s complex network may be highly geographically dispersed and functionally diverse, they are likely to be highly interconnected. Rosenau describes the implication of the equivalent phenomenon in the international system as it becoming “dense” with causal layers.\(^97\) This support’s Kilcullen’s assertion that actors will be highly nested, and that attacking a malign actor will inevitably lead to harm to other actors. Second, as each actor can be represented by its own subsystem, the complex network is actually a series of tightly-interconnected and nested subsystems. Network theory tells us that shocks can cascade up and down these subsystems. So, in such a network, it is unlikely that a military intervention aimed at an opposing military actor would not reverberate across the wider network in unforeseen ways. Conversely, it may be possible to affect an opposing military actor through a non-military intervention. So whether to mitigate the undesired effect of a military action, or to achieve a military effect through a non-military action, network theory would support the claim that an integrated response is more likely to be required. Third, a scale-free structure will result in the complex network self-organising to a critical state, causing it to experience fluctuations of all scales conforming to
a power-law distribution. The fat-tail of this distribution makes the high-magnitude but low-probability fluctuations statistically significant and will cause change in the network to be perceived as a series of punctuated equilibriums.

Implications for the strategic level of war

What, then, are the implications of increasing complexity for the strategic level of war? We can draw three key deductions.

First, war conceptualised as a complex network undermines bipolar models of war. Clausewitz famously defined war as “nothing but a duel on a larger scale”, and that a “picture of it as a whole” could be derived from the image of two wrestlers. The contention of this paper is that, even though at its heart a war may be a confrontation between two actors, there remain third-party actors who have strategic impact on its outcome. A picture of war as only two actors is, therefore, an incomplete picture. The inference here is not that all Clausewitz’s insights should be dispensed with. We have already seen how Clausewitz’s thinking on policy’s restraint of war’s escalation is entirely coherent with complexity theory. Alan Beyerchen similarly remarks that Clausewitz’s understanding of the interaction inherent in war makes it “consistent with our current understanding of nonlinear dynamics.” Clausewitz was clearly aware of the complexity caused by multiple actors, not least by subdividing each actor into a trinity of sub-actors. But Clausewitz’s suggestion that ‘a picture of war as a whole’ can be gleaned from a duel is, to paraphrase Korzybski, no longer as ‘useful’ a map of war as one of it as a complex network. Furthermore, as we shall see in Chapter Four, such a reductionist conceptualisation of war leads to overly simplistic analytical methods for campaign planning.

Second, an increasingly complex world will increasingly self-organise far from equilibrium. Our models of the world need to be updated to reflect this condition. Economics has had to do this in the wake of the 2008 financial crisis. Neoclassical economics, for some time the mainstream of economic thinking, encourages a micro-economic approach, where the behaviour of a financial system is understood from the actions of its individual agents. In this sense it is inherently reductionist and does not account for emergent behaviour in the complex financial system. Crucially, it also assumes markets self-organise towards equilibrium and, therefore, struggles to explain very large fluctuations like the recent financial crisis. Similarly, structural neorealism, the mainstream international relations approach for the latter half of the twentieth century, is also an essentially equilibrium-based theory. It assumes states operating amid anarchy self-organise to a form of equilibrium (a balance of power). Kenneth Waltz, perhaps the father of structural neorealism, drew extensively on the cybernetics and system theorists we have characterised as the antecedents of complexity theory. He, therefore, criticises reductionist theorists for not accounting for emergent system behaviours, a view he almost certainly gained via Bertalanffy and which remains coherent with modern complexity theory. However, he also agreed with Bertalanffy’s belief that open systems would self-organise towards equilibrium. His theory of international politics “predicts a strong tendency toward balance in the system”. But we have already seen that Bertalanffy’s premise that systems self-organise towards equilibrium has been over-turned by complexity theory, where the consensus view is now that complex systems self-organise into a far from equilibrium state of criticality. In essence, the mainstream theories of both economics and international relations are based on outdated systems theory and need updating.

Third, a world which self-organises to a state of criticality will unfold as a series of punctuated equilibriums. The majority of change and upheaval in the international system will come about through rare but statistically significant upheavals. This does not guarantee an increase in the likelihood of war. Violence is only one potential response to such upheavals (although a consistent response in the past). It does, however, challenge conventional risk management methodology. Mandelbrot, writing with Nassim Nicholas Taleb, has argued that the impact of ‘outlier’ events demands a different approach to risk. A traditional risk management approach applies a ‘Gaussian’ risk distribution, one where the consequences of outlier events are mild and therefore have little bearing. However, a system conforming to a fat-tailed power-law distribution experiences outlier events so dramatic that it warrants placing them as the starting point of risk mitigation strategies. With the exception of the nuclear deterrent, the risk
management methodology within the United Kingdom National Security Strategy is arguably more Gaussian than fat-tailed. It stratifies national security risks across three tiers based on a combination of their likelihood and probability. Thus higher magnitude threats have their overall rating severely diminished by their low probability of occurrence and generally get placed at tier three (examples are a conventional attack on another NATO member and disruption to international resource supplies). Resources are prioritised against the highest tier risks and this logic is carried through into the United Kingdom Strategic Defence and Security Review. Higher-magnitude risks are mitigated through either the nuclear deterrent (the one exception of resourcing a low-probability, high-magnitude threat) or an assumption that conventional capabilities can be regenerated given “sufficient strategic notice.” But as we have seen, a complex international system self-organised to a critical state will not give strategic notice, meaning that the United Kingdom’s response to high magnitude threats is potentially limited to its nuclear deterrent, which offers a drastic and somewhat limited range of responses to decision-makers. Complexity theory would therefore suggest that in a complex world defence prioritisation should be more heavily weighted towards low-probability but high-magnitude events.

In the next chapter we will consider the implications of increased complexity for the operational level of war.

CHAPTER 4 – COMPLEXITY AND THE OPERATIONAL LEVEL OF WAR

This chapter considers complexity and the operational level of war. It first reviews two previous doctrinal attempts to deal with the increased complexity of war. It then employs recent complex network theory to investigate the implications of designing campaigns within a model of war which resembles a complex network. This is done from three perspectives: exerting control within a complex network, understanding a complex network, and predicting effects within a complex network.

Western military responses to the increasing complexity of war

As the ‘new war’ thesis developed amongst academics in the 1990s, so military doctrine writers investigated how to successfully employ military capability in such wars. Tim Bird concludes that during this time doctrine development became dominated by “the problems associated with the planning and use of force in complex scenarios in which the military component may well be necessary, but is unlikely to be sufficient to achieve strategic objectives”. Military doctrine often sought insights from the various strands of complexity theory emerging during that time, unsurprisingly given the traditionally synergistic relationship between the military and science. The use of complexity theory in military thinking was greatest in the United States, again perhaps unsurprisingly given its pre-eminence in both military affairs and complexity science. Examples of this include a 1996 National Defense University symposium which brought together leading proponents from each field, and the direction of the Commander of the Marine Corps Combat Development Command to formally incorporate the “implications of chaos and complexity theory” into Marine Corps doctrinal manuals. Two of the most prominent strands of military thinking over this time were Network Centric Warfare (NCW) and Effects Based Operations (EBO).

NCW was first articulated in an article by Vice Admiral Arthur Cebrowski in 1998. Clearly influenced by the work of Alvin Toffler discussed in Chapter One, Cebrowski’s article stated “nations make war the same way they make wealth”, a direct (but unreferenced) lift from Toffler’s *War and Anti-War*. Cebrowski described NCW as a shift from a warfighting style characterised by attrition to one characterised by “the new concepts of speed of command and self-synchronization.” Cebrowski’s use of complexity theory can be seen both in the idea of self-synchronization (his article states “complexity theory tells us that such enterprises organise from the bottom-up”) and in his use of the prototypical complex network, an ecosystem, in his vision that the United States military become “a more powerful warfighting ecosystem.” But Cebrowski’s conception of an ecosystem was not of the United States military as one actor within a multi-actor complex network (the model we derived in Chapter Three) but simply of the various elements of the United States military, its intelligence, surveillance, reconnaissance and precision.
firepower systems. This conception was, according to Arent Arntzen and Tor Olav Grøtan, a “theory of war that does not encompass any significant notion of an enemy.” Peter Mansoor similarly concludes that NCW created a force perfectly suited to fighting itself.

But Cebrowski’s influence would prove far from academic as he subsequently founded the United States Office of Force Transformation in 2002, implementing military transformation under Secretary of Defence Donald Rumsfeld. Emboldened by the successful combination of information and precision weapons in the 2001 Afghan War, Rumsfeld set about enacting a transformation programme which focussed on winning interstate conflicts but virtually disregarded all intrastate scenarios. It prioritised air and naval capability to the detriment of ground capability, generating forces that Olof Kronvall concludes were “good at war winning but not effective at peace winning.” This became apparent when the United States force structure proved inadequate to deal with the insurgencies in both Iraq and Afghanistan. Other officers (such as Generals David Petraeus and James Mattis) successfully lobbied for greater emphasis to be placed on counterinsurgency operations, which resulted in the 2006 National Security Strategy recognising counterinsurgency on a par with interstate warfighting. While some authors suggest this was merely a continuation of the transformation programme, it was, in fact, a more fundamental change of policy, with the Office of Force Transformation closing in 2006.

While NCW was primarily an ideology to guide force development, EBO was a doctrine for the employment of force. It was designed to “deal with the complexity of modern missions and their implications for linking military actions and operations to diplomatic, information, economic, and social actions.” It became an official element of United States doctrine in 2001, but had its roots in United States Air Force thinking going back to the 1990s, thinking typified by the work of Colonel John Warden. Warden drew on systems theory and argued that the enemy should be conceptualised as a system. He proposed a five ring model of the enemy system and argued for a ‘strategic’ form of warfare, which sought paralysis of the enemy through the targeting of the inner rings (the leadership and essential services of the enemy regime) rather than merely targeting its fielded military (which formed the outer ring). Within his paper Warden argued:

*The more complex a system, the more precarious its maintenance tends to be and the more likely that injections of energy in the wrong places will speed its natural movement towards disorder and perhaps even chaos.*

Parallels have been drawn between Warden’s model and complex adaptive system theory. Warden himself admits to being influenced by the related concepts of chaos theory. By the time Warden’s Air Force thinking had been translated into Joint Forces doctrine in 2001, it had become a method for employing all levers of national power to achieve effects. In this respect it was a logical response to the increased diversity of actors present in contemporary warfare. However, it had also developed a set of highly quantitative tools, such as Operational Net Assessment (ONA), to identify the “nodes and links” which, if targeted, would create desired effects. Such an approach had two significant failings which ultimately prevented EBO from succeeding. First, ONA created such a complex staff-led planning process that, when employed on a British exercise, it proved to be a source of chaos in its own right. NATO therefore chose to adopt a toned-down version of EBO called Effects Based Approach to Operations (EBAO) which was more a “way of thinking” than a pseudoscience. Second was the suggestion that first-, second- or even third-order consequences could be predicted during any planning process. This was the main reason why the United States Joint Forces Commander, General James Mattis, very publically excluded it from all forms of United States doctrine in 2008. In a letter to his subordinates Mattis’ argued that EBO assumed “a level of unachievable predictability”, and was “too prescriptive and overengineered”. Mattis logic for denouncing EBO was also steeped in complexity theory, arguing that “chaos makes war a complex adaptive system, rather than a closed or equilibrium-based system.”

NCW and EBO were both attempts to manage the complexity of contemporary war. Both drew on complexity theory to some degree and both had their merits. But, ultimately, both proved to be incomplete conceptions of war. NCW focused on friendly forces to the exclusion of any consideration of an enemy. It attempted to fuse information technology and precision weapons to produce a force better
than any other. It failed to consider that adversaries may adapt in a way which could make such capability largely irrelevant. The highly quantitative form of EBO, adopted and then denounced by the United States military, focused on enemy forces to the exclusion of considering its interaction with a friendly force. It assumed solutions could be engineered in a predictable way, but failed to account for the complexity and unpredictability that interaction brings to war. Each, in their own way, failed to deal with warfare in the form which we identified in Chapter Three, namely as a complex network made up of a greater diversity of actors than just friendly and enemy forces, which are all highly interconnected and nested within one another.

**Contemporary British military doctrine**

British doctrine also shows evidence of complexity theory’s insights. For instance, JDP-01 *Campaigning* urges commanders to “consider the whole situation and recognise that it is complex, adaptive, nonlinear, and to a certain extent unpredictable.” Contemporaneous British thinking has evolved from EBAO to one centred on “influence.” Influence is defined as the “power or ability to affect someone’s beliefs or actions” and is achieved through Joint Action, a term to describe the four forms of military capability or activity (fires, information activity, manoeuvre and outreach), which collectively can be employed to “affect an actor’s will, understanding and capability, and the cohesion between them to achieve influence.” These four activities encompass both physical and psychological activity, and places strategic communication at the heart of military strategy. Much of the EBAO approach can still be found in the Joint Action model. It involves applying both kinetic and non-kinetic effect to achieve objectives. The process by which a Joint Force Commander translates his assigned campaign objectives into decisive conditions, supporting effects and ultimately actionable tasks is campaign planning. Important within campaign planning is the concept of “Centre of Gravity”. Whilst their identification is stated as “not mandatory”, *Campaign Planning* stresses that all parties in a conflict are likely to have one, that their “analysis will allow the progression of the campaign to be defined, as objectives or decisive conditions are identified and sequenced in different courses of action”. Ultimately, they remain an important analytical method by which a Joint Commander arrives at a theory of change. A question we will address later is whether increased complexity makes their existence less likely and, even if they do exist, whether they remain the most effective method for arriving at a theory of change when designing campaigns in a complex network.

The remainder of this chapter explores understanding, predicting outcomes and exerting control within a complex network. It works backwards from the problem of control. If we had perfect understanding of a complex network, how would we best control it? It then considers how we can best understand a complex network. Finally, it considers whether we can ever hope to predict outcomes in a complex network.

**Controlling complex networks**

Most recently, the network theory research agenda has focused on the controllability of complex networks. By assimilating elements of control theory (a mathematical branch of engineering), Yang-Yu Liu and others introduced the idea in 2012 of “control centrality”, the extent to which any single node can control the whole network. Their work showed that, in some complex networks, the most powerful node can control only a fraction of the entire system. It also demonstrated that control centrality was mainly determined by a network’s degree distribution (a topographical feature). These findings have two implications. First, if even the most powerful node can control only a fraction of the network then it seems likely that power will be highly diffused within complex networks. Second, if controllability is determined by topography, then an understanding of the network could provide an understanding of its level of controllability. In 2013 Tao Jia and others developed this work by considering control configurations for complex networks. Their work identified that most complex networks have multiple control configurations, that is more than one combination of nodes by which the system can be controlled. By analysing all possible combinations they were able to categorise each node by the roles it played in various configurations. A node which appears in all possible control configurations is categorised as critical, a node which features in some but not all configurations is categorised as
intermittent, and a node which never features is categorised as redundant. This led to the further discovery of two distinct control modes for complex networks: centralised and decentralised control (see Figure 5).

![Centralised Control](image1)

![Distributed Control](image2)

Figure 5 – Bimodal control of complex networks

Centralised control is where network control can be achieved through a small fraction of nodes, whereas in distributed control it is necessary to control a large amount of nodes, possibly more than 90 percent. The finding of Jia was that, as the average degree distribution of the system increases, so the system experiences a bifurcation in control (see Figure 6). The control mode which a system adopts depends on its degree asymmetry (another feature of the network’s topology). While the control mode could be altered by small perturbations (for instance in Figure 5 the control mode is changed by reversing a single interaction highlighted on the left of the network) Jia concluded that most real networks were “destined” for either centralised or distributed control, and that to alter this mode was either impossible or infeasible.

![Bifurcation in control modes of complex networks](image3)

Figure 6 – Bifurcation in control modes of complex networks

While Jia’s findings have yet to be applied to war, it is clear that they could have significant implications for a revised conception of war as a complex network. We do not know the average degree distribution (average shortest path between any two nodes in the network) of a prototypical war. We can surmise that the new wars have two contradictory effects on average degree distribution: increasing the number of actors increases it, but increasing connectivity shortens it. But if we assume war is an inherently human phenomenon, then we can assume it has an average degree distribution similar to a typical social network, which empirical study has shown to be around six. Thus, from Figure 6, we can see that around 50 percent of nodes need to be controlled to achieve network control. But a note of caution. The
notion of control discussed in network theory is absolute. However, it is surely unrealistic to expect absolute control in war where there will undoubtedly be a competition for control (a phenomenon not yet investigated by network theory). So while we would like absolute control, in an adversarial situation we will generally need to be satisfied with achieving sufficient control so that we can achieve our ends and avoid those of opposing actors. A useful way then to conceptualise campaigning in our revised model of war is as a ‘competition for network control’.

Clausewitz suggests that “…the first, the supreme, the most far-reaching act of judgement that the statesman and commander have to make is to establish the kind of war on which they are embarking.” Within a competition for network control this equates to establishing the degree of network control that is required. If our aim is simply to defeat an adversary then, arguably, we do not need to control the whole network, we just need to collapse the enemy’s system. But if our aim is more comprehensive, such as in stabilisation, then our objective does move more towards absolute control. So we can potentially delineate two meta-forms of warfare: systemic collapse or network control. Let us consider each in turn.

In the case of pursuing systemic collapse we could choose to map the enemy as a discrete military system, and then calculate the most efficient way to collapse it (as Warden’s ‘Five Rings’ model does). In this case the notion of centre of gravity retains utility as it could help in identifying the vulnerabilities which will lead to collapse. A purely military response may also prove sufficient to achieve our aims in such a case. If, however, we map the enemy military system within the context of its wider network we may arrive at a different defeat mechanism. The enemy system will invariably have connections and dependencies with other systems (for instance criminal, illicit, and financial systems), which could be either a source of resilience or vulnerability. By only conceptualising the enemy as a discrete system we may either underestimate an enemy source of strength or miss a route of attack. Kilcullen explains how the Shabaab insurgency in Somalia has links with piracy, European organised crime, clans, criminal gangs in Mogadishu, humanitarian organisations and legitimate businesses. Any of these connections could serve as a source of strength for Shabaab, which if only conceptualised as a discrete military system, would be ignored (and therefore their overall strength underestimated). Alternatively, any could serve as a source of vulnerability which, if targeted, could offer a more efficient defeat mechanism than a purely military one. So even when pursuing systemic collapse it would seem logical to still do this within a network context. And when pursuing systemic collapse within a network context it becomes less likely that a purely military response will be sufficient to achieve our aims, and that centre of gravity of analysis will arrive at the most efficient defeat mechanism.

But what if we are attempting to achieve network control? We have assumed earlier that 50 percent of nodes are required to achieve absolute control of the network (the other 50 percent being redundant). In a bipolar competition for network control, assuming that all nodes exert equal control, we could further assume that control of 26 percent of nodes will provide sufficient control (leaving our opponent with only 24 percent of controllable nodes). The balance of control tips in our favour, allowing us to control the network more than our opponent. The 26 percent of nodes that we need to establish sufficient control are likely to exist across a variety of systems (both military and non-military). It therefore seems intuitively unlikely that the network can be controlled through purely military means, and a purely military response is likely to be insufficient to achieve our aims. In addition, when controllability is so diffused, conducting a centre of gravity analysis on only one or a few actors is unlikely to be most effective analytical tool for deriving the theory of change required for the campaign.

So complexity has potentially highly significant implications for exerting control at the operational level. It supports the new war argument that purely military responses are unlikely to be sufficient, except in the limited case of systemic collapse of a military actor. It also calls into question the analytical validity of current campaign planning methods.
Understanding complex networks

To this point our discussion has assumed perfect understanding of the network, a clearly unrealistic assumption. So the question remains, can we ever hope to understand a complex network sufficiently to make use of (the otherwise purely academic) insights above? While intuitively it may feel that increased complexity within a system would hinder our ability to model it accurately, there is both academic and practical evidence that the reverse is the case. Increased connectivity could provide unprecedented opportunities to map and make sense of war.

From an academic perspective Liu and others have considered the degree to which complex networks are observable, that is whether “we can reconstruct the system's complete internal state from its outputs.” Their work sought to identify the minimum number of nodes (the “minimum sensor set”) which contained sufficient information to reconstruct the system's complete internal state, thus making it observable. Their theoretical finding was that observing a single sensor node in each of the network’s core sub-systems (its “strongly connected components – SCC”) was sufficient to infer the state of the entire network. As an example, in Figure 7, the state of the entire 17 node system can be accurately inferred from observing three identified sensor nodes.

The implication of this finding is the possibility that we can gain an accurate understanding of war’s complex network through selecting and observing a relatively small proportion of sensor nodes within it. Importantly for achieving this, a common contention is that the increasing availability of global data reduces the opacity of human dynamics. For instance, Barabási claims “currently social systems offer some of the best mapped datasets on the dynamics of any complex system, when the action of each component (individual) can be followed in ultimate detail.” Similarly, and from a practitioner’s perspective, Kilcullen observes that “real-time crisis mapping” (a combination of field-data and open-source data) can “produce an incredibly detailed and workably accurate map in near-real time as a crisis unfolds.”

This increase in the available data of social systems is so significant that Global Pulse, the United Nations laboratory for data innovation, has termed it a “global data revolution.” Possible sources of such data are: data exhaust (passive transactional data such as purchase and web searches), online information (such as blogs and twitter), physical sensors (such as satellite imagery or traffic patterns) and crowd-sourced data (mobile-phone surveys or user-generated maps). This data can vary between open-source, proprietary and classified. An example of the breadth of insights which can be mined from such datasets is the 2013 Data for Development Challenge, in which over 80 research projects were undertaken on a massive anonymised dataset of mobile-phone calls in the Ivory Coast (an example of normally proprietary data). The results showed that from mobile-phone data it was possible to: map where people live and work, chart human mobility, identify how far the influence of areas extends, and detect unusual and significant events. Data-mining has already become a standard feature of military
activity, through the use of data-mining tools such as Palentir which has identified terrorist and insurgent networks, and even purportedly had a role in locating Osama Bin Laden. The contention of this paper is that the data revolution combined with the insights of complex network theory offer us ways of mapping more accurately complex networks, and understanding how best to observe and control them.

**Predicting in a complex network.**

As we have already noted, the nonlinear dynamic interactions and emergent system-level behaviours of complex networks cause them to be highly unpredictable. And yet, a key part of designing and executing campaigns is predicting the likely effect of an intervention. So any insight that complexity theory can offer in increasing accuracy of predictions is likely to be valuable.

Bak introduces the concept of signal and noise in *How Nature Works*. This is the contention that, amongst the random noise of any complex network there is a signal which underpins the patterns of the system’s fluctuations. The implication for us is that there could be a signal within the complex network of war which can offer insights into its dynamics. But a side-effect of the data revolution is an exponential increase in the amount of noisy-data, data which only pollutes this true signal. Nate Silver explores this concept further in his book on prediction, *The Signal and the Noise*. Silver observes that some disciplines associated with complex phenomenon have achieved relatively good levels of prediction (such as weather prediction), whereas others have remained exceptionally poor (such as earthquake prediction). The difference lies in the amount of observable data, and the degree to which that data is either noisy or approaching pure signal. Furthermore, Silver stresses the benefits of a Bayesian approach which “encourages us to hold a large number of hypotheses in our head at once, to think about them probabilistically, and to update them frequently when we come across new information”, an approach, he stresses, as being particularly appropriate for the realm of national security analysis.

An example of signal and noise at work in a complex network can be seen in the financial system. Market data has been found to contain statistically significant signals (for instance interest rates and trading volumes) which could be used to predict future share prices. Like war, the financial system is competitive and, therefore, such signals are often arbitraged away meaning that, while some signals may persist for months, others last only days. The parallels between the financial system and our complex network model of war seem stark. Eric Beinhocker equates financial markets to an ecosystem, and observes that “the complex, nonlinear dynamics of the markets meant that new signals were constantly being created even as old signals faded due to arbitrage”. Social Intelligence, a division of PA Consulting, claim that the mass of available open-source social data makes groups behaviours highly predictable, a fact which can be exploited to map consumer needs and predict new trends.

So the contention of this paper is that there will exist, amongst the wealth of noisy-data emanating from a complex war network, signals which can aid prediction. And once again the latest complex network theory could offer insights which assist this process. Baruch Barzel and Barabási have identified a method to silence noise (by removing the effect of indirect interactions) which can improve insights about a system’s interactions and its dynamical mechanisms. Furthermore, Barzel and Barabási have also identified improved ways to predict a complex network’s response to perturbations, specifically how a perturbation in one part of the network may propagate and cascade through other parts. They suggest that a networks response to a perturbation is driven by just a small number of universal characteristics, meaning that measurement of those characteristics could improve predictions of how the network will respond to an intervention.

**Implications for the operational level of war**

The main implication for campaigning for a model of war, which resembles more closely a complex network than a duel is relatively obvious: campaign planning and execution will need to better understand the dynamics and structures of networks in order to make successful interventions. It demands a mind-set change away from dyadic, balance of power confrontations to one based on a competition for network control. Equally, campaign planning should move away from centre of gravity
analysis as the default method of arriving at a theory of change, and investigate new methods based on achieving sufficient control of complex networks where power is diffused across many actors.

CHAPTER 5 – CONCLUSION

This paper began with the observation from doctrine that war is becoming increasingly complex. Our employment of complexity theory supports that claim. Globalisation and the information age have increased the number and diversity of actors who influence war, and intensified the extent of their interconnectedness. So doing they have made war a more complex phenomenon. And yet there is also evidence that military thinking has not yet understood the full implications of increased complexity. At the strategic and operational levels of war, the universal structures and dynamics of complex systems challenge many of the assumptions and precepts of conventional thinking. Be it the way that states manage national security risks, or the way campaign planning derives a theory of change, complexity theory would question whether existing theories are optimised for coping with complexity.

The main recommendation of this paper is, therefore, that military practitioners and academics seek to understand the structures and dynamics that underpin complexity. Gaining a comprehensive theoretical understanding of complexity as a phenomenon becomes a sensible, if not essential, aim. Complexity theory can offer valuable insights into understanding the complexity inherent in all levels of war. This is perhaps the reason why American military thinking has called on complexity theory for over two decades, and why many of the recent network theory papers cited in this paper continue to be funded by the United States military. Absorbing complexity theory into military thinking will not guarantee success (as we saw with NCW), but discounting its insights may well ensure failure.

Not understanding complexity will see us continue to conceptualise war as a duel, and to follow all of the reductionist, bipolar methodologies that flow from such a conceptualisation. Thinking in duels in a world which resembles a network will cause multiple errors. We will overestimate the level of control that can be achieved through targeting a single actor. We will underestimate the unintended consequences of military actions. We will fail to see routes of indirect attack to adversaries that exist through a wider network. Understanding war as a complex network is the conceptual point of departure for designing more effective campaigns. It forces us to consider operations as a competition for network control.

In a competition for network control a successful actor will be one who understands the system, and where to leverage control within it. Network theory tells us that conducting a centre of gravity analysis on a single actor will rarely be an effective analytical method for deriving a campaign’s theory of change. In a purely bipolar military confrontation it could be useful for achieving the systemic collapse of an enemy. But in the more complex situations, which characterise contemporary war, it is overly simplistic, reductionist and incomplete. Instead, new analytical methods need to be established which better accommodate the diffused nature of control in complex networks.

The importance of data, be it open-source, proprietary or classified, is clear. In the future we may designate a dataset as vital as we have in the past terrain. Mining, buying, hacking or fighting for that data could then become a priority for a Joint Commander. Equally, the pollution of datasets could become the new means of military deception. The military and academia should investigate the potential implications of the data age as vehemently as it did the information age in the 1990s. Social Intelligence claim that “companies that are skilled in extracting intelligence from consumer data and have the ability to react fast to embryonic trends will be the victors.” The same could also be true in war.

The contention of this paper is not that we can overcome the unpredictability of war through data-mining and applying algorithms. The experience of EBO shows us that overly-engineered and quantitative methods do not succeed. But the insights of complexity theory could allow us to better observe and understand the structure and dynamics underpinning war’s complex network. In turn, we may plan more effective campaigns, gain more accurate predictive foresight and achieve better
situational understanding. Complexity is a force whose silent dynamics have the capacity to defeat us well before we ever confront an adversary on the battlefield. We cannot defeat complexity, but if we are to succeed in war, we must learn to harness it to our advantage.

Endnotes
1 DCDC, British Defence Doctrine, Joint Doctrine Publication 0-01, 4th ed. (Shrivenham: DCDC, 2011), iii.
2 Clausewitz identifies violence, the play of chance and subordination to policy as defining the nature of warfare. Carl Von Clausewitz. On War. (New York: Oxford University Press, 2006), 30. In British Defence Doctrine complexity (or its derivatives) is mentioned 25 times compared to violence 13 times, chance nine times and policy 40 times. JDP 0-01.
3 “consider the whole situation and recognise that it is complex”. UK. DCDC, Campaigning, Joint Doctrine Publication 01, 2nd ed. (Shrivenham: DCDC, 2008), 3-5.
4 UK. DCDC, Global Strategic Trends, 4th ed. (Shrivenham: DCDC, 2010).
6 Clausewitz, On War, 82.
12 Clausewitz, On War, 13.
13 For an excellent introduction to critical realism see, Heikki Patomo. After International Relations: Critical Realism and the (re)construction of World Politics. (London: Routledge, 2002).
17 Vish Rennen and Pim Martens, The Globalisation Timeline. Integrated Assessment 4, no.3. (Maastricht: International Centre for Integrative Studies, 2003), 140.
19 For instance, a third of the world’s population has access to the internet, mobile-phone penetration in the developed world reached 79 percent in 2011, and levels of mobile-internet access is overtaking levels of domestic electricity in South and Southeast Asia, sub-Saharan Africa and the Middle East. John Pollock, “People Power 2.0: How Civilians Helped Win the Libyan Information War,” MIT Technology Review, (May/June 2012), 71. http://www.technologyreview.com/featuredstory/427640/people-power-20/
20 For instance, Susan Strange argues that financial, industrial and trade forces have seen a decline in the authority of states and a growing diffusion of power to other actors. Susan Strange, Power Diffused: State and Non-state Authority in the World Economy. (New York: Cambridge University Press, 1996), 4.
21 Toffler predicted that a second ‘industrial wave’, which had replaced a first ‘agrarian’ wave, was now being overtaken by a third wave. His analysis of how this third wave would diminish the role of the nation state, increase the importance of information, and alter economies were all prescient predictions of globalisation’s impact. He did not give the third wave a single title, but one of the terms he used (the information age) became most associated with it. Alvin Toffler, The Third Wave. (New York: Morrow, 1980).
30 Ibid., 238.
33 Simpson, War From the Ground Up, 74
34 Coker, War in an Age of Risk, 177.
35 Ibid., 172.
37 Kaldor, New and Old Wars, 149.
39 Ibid., 114-116.
40 Data extracted from Ben-Yehuda, “War Complexity and Outcomes”, 136-140.
...
“The political object – the original motive for the war – will thus determine the military objective to be reached and the amount of effort it requires.” Clausewitz, On War, 20.


Cederman, “Modeling the Size of Wars,” 144.


Rosenau, Turbulence in World Politics, 9.

Clausewitz, On War, 13.


Rosenau, Turbulence in World Politics, 9.

Clausewitz, On War, 13.


Clausewitz, On War, 30-31.


Fleming, “New or Old Wars?” 228.


Waltz, Theory of International Politics, 56 and 72.

Waltz, Theory of International Politics, 128.


A Strong Britain in an Age of Uncertainty: The National Security Strategy, Cm. 7953 (2010), Annex A.

Ibid., 3.14.

Ibid., 0.17.


In an excellent exploration of the role of nonlinear science in contemporary warfare, Sean Lawson claims that there were “hundreds” of attempts to assimilate complexity thinking into military affairs during the 1990s. Sean T. Lawson, Non-linear Science and Warfare: Chaos, Complexity and the US Military in the Information Age. (Milton Park: Routledge, 2014), 16.


Ibid., 32.

Ibid.


Lawson, Nonlinear Science and Warfare, 6.


Kronvall, “Transformation: The Key to Victory?” 262.

David S. Alberts and Richard E. Hayes, Power to the Edge: Command... Control... in the Information Age. (Washington, DC: CCRP 2003), 1.


132 Ibid., 46.


134 John Warden, email correspondence with author, April 10, 2014.


137 Ibid., 793


139 Ibid., 21

140 JDP 01, 3-5.

141 Alexander Alderson presents the shift in military thinking towards influence as being an evolutionary path which includes Liddell Hart’s Indirect Approach, the Bagnall reform-inspired Maneuvrist Approach and EBAO. Alexander Alderson, “Influence, the Indirect Approach and Manoeuvre,” RUSI Journal 157, no.1 (2011): 36.


143 JDP 01, 3-13/14.


145 JDP 01, 3-10.

146 “A characteristic, capability, or influence from which a nation, an alliance, a military force or other civil or militia grouping draws its freedom of action, physical strength, cohesion of will to fight.” JDP 5, 2-17/18.

147 Ibid.

148 Specifically, control centrality is the “dimension of the controllable subspace or size of the controllable subsystem when we control that node only”. Yang-Yu Liu, Jean-Jacques Slotine and Albert-László Barabási, “Control Centrality and Hierarchical Structure in Complex Networks,” PLoS One 7, no.9 (2012), 1.

149 Ibid., 2.

150 Ibid.


152 Ibid., 2.

153 Adapted from ibid., 4.

154 Ibid., 2.

155 Ibid., 4.

156 Adapted from ibid., 3.

157 Even for social networks with as many as 6 billion individuals, Albert, Jeong and Barabási, “Error and Attack Tolerance of Complex Networks,” 380.


159 A concept in part inspired by the theory of competitive control discussed in Kilcullen, Out of the Mountains, 114.

160 Clausewitz, On War, 30.

161 Clausewitz, On War, 30.

162 Doctrine suggest goals can be as ambitious as, “changing the socio-political and economic dynamics of a society”. JDP 01, 3-22.


164 Ibid., 5.

165 Albert-László Barabási, “Human Dynamics,” Barabáslab,
http://www3.nd.edu/~networks/HumanDynamics_20Oct05/human_dynamics.htm

166 Kilcullen, Out of the Mountains, 253-255.


168 Global Pulse, “Annual Report 2013,” Global Pulse,


173 Silver, The Signal and the Noise, 444.


175 Ibid., 392.
179 Ibid., 679.
180 As examples, the Barzel and Barabási papers cited in Chapter Four were partly funded by the US Army Research Laboratory, the Office of Naval Research and the Defense Threat Reduction Agency. Barzel and Barabási, “Universality in Network Dynamics,” 681.
181 Social Intelligence, Seven Billion Faces of Social Media, 5.
What are the impediments to the creation of an independent Kurdish state?

LT COL J AINLEY RE

Introduction

The Kurds are a stateless nation of around 25-30 million people that are associated with a large swath of territory that spans the borders of Iran, Turkey, Iraq and Syria. They “lost their one great opportunity for statehood” between 1918 and 1925, when new states were being created out of the former Ottoman Empire, and have since been incorporated into those four states. This has in turn fuelled Kurdish nationalism. Yet “some of the most important developments in modern Kurdish history” have occurred in recent years. The start point can be traced to the fateful decision by Saddam Hussein to invade Kuwait in 1990, which sparked the chain of events that led, via the 2003 invasion of Iraq, to the Kurdistan region gaining formal status within a federal Iraqi state in 2005. Turkey’s journey towards EU accession that began in 1999, followed by the election of the Justice and Development party (AKP) in 2002, has led to a series of reforms, openings and a peace process that, although often criticized, has seen “enormous progress” made with regards to Turkey’s Kurdish question. Finally, Syria’s Kurds have experienced de facto autonomy since mid-2012 amid the chaos of the Syrian civil war. In light of these developments, therefore, this paper will seek to identify the impediments to the creation of an independent Kurdish state. An understanding of these impediments will help defence planners understand why “Kurdish nationalism remains a continuing and leading factor of instability in the geostrategically important Middle East”.

This study will be conducted over three chapters. Chapter one will examine the Kurdish nation itself, in order to identify the internal impediments. It will show that differences in religion and language exist but, for the most part, are inconsequential. However, tribalism and political disunity has repeatedly undermined attempts by the Kurds to better their position. That, combined with a number of factors associated with the territory occupied by the Kurds, presents a considerable barrier to the attainment of Kurdish independence.

Chapter two will examine the international dimension. It will start by examining the historical position of the Kurds in relation to the other interests of the major powers. It will then investigate whether the international system is likely to support the creation of an independent Kurdish state. It will show that only in exceptional circumstances, and with considerable international support, would the international system prioritise self-determination over the preservation of territorial integrity. Therefore, having determined earlier in the chapter that Kurdish interests are often subordinate to the other interests of the major powers, chapter two will conclude that the Kurds are unlikely to receive the level of support necessary to achieve statehood.

Chapter three will consider the situation of the Kurds within Turkey, Iraq, Iran and Syria. The start point will be that all those states would oppose the creation of an independent Kurdish state and, for that reason, as well as for the reasons identified in the first two chapters, Kurdish independence is unlikely. The impediments identified in this chapter will therefore not only relate to Kurdish independence, but also to lesser Kurdish aspirations within each state.

The study will conclude that, despite recent events resulting in an improved position for the Kurds within some states, independence remains very unlikely. Kurdish disunity and the tendency of the international system to prioritise the preservation of territorial integrity over claims of self-determination are the two main impediments to that goal. At the state level, additional impediments to independence or other Kurdish aspirations are the influence of external actors, notably Turkey, and the inability of state governments to make reforms.
Chapter One – The Kurds

The Kurds are of Indo-European origin. They are frequently described as descendants of the Medis or Gutis tribes. However, it is probably more accurate to describe them as “an amalgam of the many groups that made their home in the Anatolia/Zagros axis, and of those who passed through on their way to elsewhere”. The Kurds can be defined by the territory with which they are associated, their tribal culture, shared history, language and religion. Yet anomalies and differences can be found under all of those headings. Moreover, those differences have yielded a variety of peoples that, to varying degrees, can be described as a Kurd. Martin van Bruinessen describes the Kurds as having “a core whose ethnic identity is unambiguously Kurdish, surrounded by more peripheral groups whose identification with Kurdish ethnicity is only one of several options”. Nevertheless, the Kurds have broadly recognised themselves as a separate people for at least the past few centuries. Most importantly, they have been recognised by others as such since at least the sixteenth century, despite the efforts of some states to deny that.

Taking an acceptance of the Kurds as a distinct grouping as a start point, this chapter will examine the territory, language, religion, tribal culture, politics and aspirations of the Kurds, in order to identify any impediments to Kurdish independence that are internal to the Kurdish nation. Before doing so, however, it is necessary to clarify the difference between a nation and a state. A state is a formally recognised entity within the international system that is most often described, using the Montevideo criteria as having “a defined territory, a settled population, effective governance and the means to enter into relations with other states”. The term nation is also commonly used to describe a state. However, in studies such as this it has a different meaning, of which there is no commonly agreed definition between scholars of nationalism. Therefore, for the purposes of this paper, Anthony Smith’s definition is used. He defines a nation as being “a named human community occupying a homeland, and having common myths and a shared history, a common public culture, a single economy, and common rights and duties for all members”. Smith further explains that a nation “is not a state, because the concept of the state relates to institutional activity, while that of the nation denotes a type of community”. It is this understanding that is used when referring to the Kurds as a nation in search of a state within this paper.

Territory

The geography and location of Kurdistan has had a significant impact on both the cohesion of the Kurdish nation and on its prospects of attaining an independent state. Mountains, in particular the Zagros range, dominate the region. They have both impeded Kurdish unity, while also becoming a prominent part of Kurdish culture. Kurdistan’s presence astride the borders and within the sovereign territory of Turkey, Iraq, Iran and Syria has presented both impediments and opportunities. States, like the Ottoman and Persian empires before them, have used Kurdish tribes or groups to resolve issues on their frontiers and to foment unrest in neighbouring states. Kurdish groups, such as the Kurdistan Workers Party (PKK), have been able to seek refuge and resources across the international borders. Economic benefits have also been reaped from smuggling and by seeking opportunities during times of international sanctions, such as in Iraq during the 1990s. Like the mountains, the presence of the international borders has, however, damaged the coherence of the Kurdish nation. “Separate histories, for at least the past [eighty] years, and different processes of socialisation have made the Kurds of each country rather different from those in the neighbouring states”.

Two other factors associated with the region of Kurdistan will impede, or at least complicate, the creation of an independent Kurdish state. First is the wealth of natural resources and fertile land contained within it. This benefits the Kurds in terms of providing a potential route towards economic independence, but it is also a reason why states are unlikely to give up that territory easily. Of particular significance is the large reserves of water contained within the Kurdish regions of Turkey and Iraq, and the oil contained within the Kurdish parts of Iraq. Second, is the small number of other ethnic groupings that reside within the Kurdistan region, which include Turcoman, Assyrians, Armenians, Yazidis, Turks, Arabs and Persians. Trapped minorities and natural resources are reasons why Turkey, Iran, Iraq, Syria, and the wider international community would likely oppose the creation of an independent Kurdish state.
Language

Language is an important part of Kurdish identity despite the existence of a number of different dialects. The two main dialects are Kurmanji and Sorani. Kurmanji is spoken in northern parts of Kurdistan, including in Turkey, Syria, and northern parts of Iran and Iraq. Sorani is spoken in Iraq and in most of the Kurdish speaking areas of Iran. These main dialects have been described as “[differing] as much from each other as English and German”. A further two dialects are also associated with the Kurdish region: Zaza, which is spoken in the north-west in vicinity of Diyarbakir; and Gurani, which is spoken in southern parts of Kurdistan. To varying degrees, all of these dialects have come under attack from states that have sought to repress or erase Kurdish identity. Yet the fact that they have more or less survived, and that the Kurds have retained a sense of common identity despite these language differences, is an indicator of the depth and resilience of Kurdish identity.

However, linguistic differences can divide the Kurds. An example is the opposition of certain Kurdish nationalists to the development of Zaza as a Kurdish language, on the grounds that it undermines the unity of the Kurdish nation. This caused some Zaza speaking Kurds to seek to establish Zaza speakers as a separate and distinct ethnic grouping. There is also a minor division that exists between the Kurmanji speaking northern parts and the Sorani speaking southern parts of Iraqi Kurdistan. For the most part, however, language is only a minor obstacle to Kurdish national unity.

Religion

Religion can divide the Kurds, yet historically it has also acted as a mobilising force that has transcended other divisions within Kurdish society. The majority of Kurds are Sunni Muslims who adhere to the Shafi’i school. A significant proportion in southern parts of Kurdistan are Twelve Shii Muslims like their Iranian and southern Iraqi neighbours. Heterodox, or to some, non-Islamic religions, including Alevi, Yezidi and Ahl-i Haqq, account for the remainder. There also exists a small number of Kurdish speaking Jews and Christians who are not usually considered as Kurdish.

The extent to which religion can be used to distinguish Kurds from non-Kurds varies. Traditionally, Kurdish Sunni Muslims were associated with the more mystical forms of Islam, in particular the Sufi orders. Both the Shafi’i school and Sufi orders therefore distinguished Sunni Kurds from their Sunni Arab and, to a lesser extent, Turk neighbours, the majority of which adhered to the Hanafi school. In Turkey, secularisation, state education and urbanisation have in part eliminated those differences. Kurds in Iraq, on the other hand, were not subjected to the same level of state interference in religious affairs. Kurdish forms of Islam in Iraq have therefore remained more distinct from mainstream Islam. The minority Kurdish religions can sometimes relate more to non-Kurds of the same religion than Kurds of other religions. For example, the Alevi Kurds in Turkey have more in common with the Alevi Turks than with non-Alevi Kurds. Similarly, the Shi’i Kurds of Khorasan in north-east Iran have remained outside of the Kurdish nationalist movement altogether.

Those competing for influence over the Kurds in Turkey have recognised the importance of religion. The PKK who, like other left wing nationalist movements, previously viewed religion as an impediment to progress, created religious groups under its umbrella in the 1990s in order to broaden its support base. Similarly, in the past decade the AKP has sought to use Islam to compete for Kurdish votes in Turkey’s south-east. Thus, for the majority, religious differences are not an impediment to the creation of a Kurdish state. In fact, it is one characteristic that can transcend other divisions within Kurdish society.

Tribe

Tribes have been, and in many parts remain, a prominent feature of Kurdish society and politics. Like religion, tribalism is viewed by nationalists as an obstacle to the achievement of their goals. History also supports that view. Tribal disunity is believed to have undermined nearly every Kurdish revolt undertaken during the nineteenth century. Moreover, many opportunities since then have been
squandered due to a lack of unity among the Kurds caused, in many cases, by tribal divisions, or rivalry between groups that are based upon tribal allegiances.\textsuperscript{38}

The reason for that disunity is that tribes have frequently prioritised their own interests above all else.\textsuperscript{39} Tribes have, therefore, been as willing to work against the Kurdish nationalist movement on behalf of the state as they are to work with it, especially when given the opportunity to strengthen their position relative to their rivals.\textsuperscript{40} The Hamidiye Cavalry, formed from Kurdish tribes in the late nineteenth century by Sultan Abdulhamid II, is one example of this.\textsuperscript{41} More recent examples include the Jash Militia, employed by the Ba'ath Party in the late 1980s to suppress the Kurds in Iraqi Kurdistan, and the ‘village guards’, employed by the Turkish State as a counter to Kurdish nationalism in eastern Turkey.\textsuperscript{42} Tribes are, therefore, also susceptible to ‘divide and rule’ type tactics that can be employed by states to undermine Kurdish unity.

Urbanisation, greater access to education and employment, and land reform are just a few of the reasons why tribalism is no longer as prevalent in Kurdish society.\textsuperscript{43} It has not, however, diminished altogether. Tribalism still features in Iraqi Kurdish politics. Kurdish political groups in Turkey and Iran are not as tied to tribal networks as those in Iraq, but tribes are still important in the rural societies of both states.\textsuperscript{44} Tribalism therefore remains a threat to Kurdish unity.

\section*{Politics}

Similarly, most Kurdish political parties will often place their own interests above all else, forge alliances with whoever can help them pursue those interests, and will seek advantage over their rivals rather than work together towards the same goals. The alliances and rivalries of the Kurdistan Democratic Party (KDP) are illustrative of this type of behaviour.

The Barzani dominated KDP has fought for primacy over the Kurdish region of Iraq since its formation in 1946, first against internal elements headed by Ibrahim Ahmad, and then his son in law, Jalal Talabani, who broke away from the KDP to form the Patriotic Union of Kurdistan (PUK) in 1975.\textsuperscript{45} Barzani represented “the more conservative and traditional, tribal wing of the KDP”, whereas Ahmad and Talabani represented a more militant and Marxist ideology.\textsuperscript{46} Both, however, sought broadly the same outcome: autonomy for Iraq’s Kurds within a democratic Iraq.\textsuperscript{47} It is worth noting the tribal, religious and linguistic origins of the two parties, in order to reinforce some of the points mentioned above. The KDP was built around a network of tribal alliances centred on the powerful Barzani family in the Kurmanji speaking northern part of Iraq around Erbil and Dohuk.\textsuperscript{48} The PUK is built upon alliances between Jalal Talabani and Kurdish leaders in the Sorani speaking southern part of Iraqi Kurdistan, centred on Sulaymaniya.\textsuperscript{49} Both belong to different Sufi orders: Barzani to the Qadiri, and Talabani to the Naqshbandi.\textsuperscript{50} Religion and language no longer divide the two parties.\textsuperscript{51} However, the KDP and PUK still rely on their respective support bases in the north and south.

Throughout its struggle for primacy, the KDP has formed a number of unlikely alliances to strengthen its position. These have included one with Iran during the 1960s, with Saddam Hussein’s government during the mid-1990s, and with Turkey during the past two decades.\textsuperscript{52} These alliances have often come at the expense of relations between the KDP and Kurdish groups within those countries and within Iraq. In fact, the KDP has gone as far as supporting Turkey in its fight against its own Kurdish nationalist movement, the PKK.\textsuperscript{53} This served to protect Barzani’s relationship with the Turkish government, while also strengthening the KDP’s position relative to its northern neighbour the PKK.\textsuperscript{54} The KDP’s rivalry with the PKK also extends towards the prominent Kurdish group in Syria, the Democratic Union Party (PYD), which is linked to the PKK and also happens to be favoured by the PUK.\textsuperscript{55} Thus, alongside its struggle for supremacy within Iraq, the KDP, the “preeminent party in modern Iraqi Kurdish history”,\textsuperscript{56} has found itself at odds with the dominant Kurdish groups in Turkey and Syria, the PKK and PYD.

Divisions within and between Kurdish political parties in Iraq are also mirrored in Iran and Syria. Iranian Kurdish political parties have been described as being “heavily fragmented”.\textsuperscript{57} Similarly, Syrian Kurdish political parties have been described as “divided”\textsuperscript{58} and “notoriously fractured”.\textsuperscript{59} Disunity among
Kurdish political parties within Turkey is not a significant factor because of the dominance of the PKK. That, however, creates its own problems. It is difficult for other Kurdish parties within Turkey to find space to put forward their agendas and to distance themselves from the PKK, an organisation that continues to be viewed as a terrorist organisation by the Turkish government.60

The failed attempt to convene a pan-Kurdish national conference in November 2013 is further illustrative of the divisions that exist between Kurdish political parties.61 To start with, broad support for such a conference indicates that most recognise a need for greater unity among the Kurds. One report stated that “[there] is some hope that all the attendees will sign a charter pledging that infighting between the Kurds will stop and that the Kurds will unite politically”.62 In Iraq, disagreements about the location and presidency of the conference then surfaced between the KDP and PUK. In Turkey, some Kurdish groups reportedly supported Barzani as head of the conference in order to counter the influence of the PKK.63 Regionally, disagreements between the KDP and PKK emerged over the nomination of delegates, and about overall control of the conference.64 In short, disunity among Kurdish political groups appears to be as damaging to the Kurdish nationalist cause now, as tribal disunity was during the nineteenth and twentieth centuries.

Exploiting the tribal and political divisions

Turkey, Iraq, Iran and Syria have been able to exploit tribal and political divisions in order to use the Kurds to unbalance their neighbours, or to contain their own Kurds. Turkey has been more inclined towards the latter than the former. Principally it has used Iraq’s Kurds, mostly the KDP, to assist in its struggle with the PKK.65 Iraq has engaged in both. It provided limited support to the Iranian Kurdish Democratic Party (KDPI) during the Iran-Iraq War,66 and has also employed ‘divide and rule’ tactics with its own Kurds. For example in 1969, the Ba’ath Party invested heavily in those areas allied to Talabani’s faction of the KDP in an attempt to divide Iraq’s Kurds.67 Iran has used Iraq’s Kurds against Baghdad, and also to keep its own Kurds under control. It did the former by supporting Barzani’s KDP during the 1970s, and also during the Iran-Iraq war.68 In 1996, Iran also provided support to Talabani’s PUK in its fight against Baghdad and the KDP.69 Throughout, Iran has been able to keep its own KDPI in check through its alliance with Iraqi Kurdish parties.70 Iran is also suspected, by Turkey, of providing support to the PKK during times when the relationship between the two has been strained, such as during the recent Syrian crisis.71 The Syrian government supported the PKK during the 1980s and 1990s during its disputes with Turkey over the Sanjaq of Alexandretta and Euphrates River water, and again most recently over Turkey’s support for the Syrian opposition.72 These many examples highlight how easily states can contain the Kurds by exploiting divisions. It is also of note that states have been able to maintain relations with neighbouring Kurds, while at the same time suppressing their own Kurdish populations.

Aspirations

Kurdish aspirations might also impose a limit on the extent to which Kurdish independence can be achieved. Martin van Bruinessen believes that “[although] most Kurdish nationalists nurture the dream of a united and independent Kurdistan”, all have now limited their ambitions to a solution within their respective states.73 This position is evident throughout the region.

Prior to his capture in the 1990s, Abdullah Ocalan, the leader of the PKK, had moderated his objective of “independence for all of Kurdistan” to a “solution within the framework of the Republic of Turkey”.74 More specifically, since 2011 Ocalan has sought ‘democratic autonomy’, a concept which is described as vague and not well understood even by Kurds.75 The Peace and Democracy Party (BDP), the current main political party within Turkey that represents Kurdish interests, similarly “demands greater autonomy from the state, including devolution and expanded cultural and political rights”.76 Furthermore, according to two studies, less than 10% of Turkey’s Kurdish population supports the creation of an independent Kurdish state.77

Notwithstanding disputes over oil and Kirkuk, the main parties of Iraq appear content with the autonomy provided by the 2005 Iraq constitution.78 That said, Gareth Stansfield does not rule out the KRG leader
aspiring to lead a “wider pan-Kurdish entity” despite Barzani’s claims to the contrary.79 Furthermore, threats of secession and independence do occasionally spill out of the ongoing disputes between Baghdad and Erbil.80 However, in contrast to Turkey, it is believed “that there is an increasing disconnect between the [KDP and PUK] and society at large”.81 During an unofficial referendum held alongside the 2005 Iraqi elections by the Kurdish Referendum Movement, 98.8% from a turnout of almost two million voted for Kurdish independence.82

Despite the differences in their methods and a lack of overall unity, the most prominent Kurdish groups within Iran, the Free Life Party of Kurdistan (PJAK), Komala and the KDPI, seek nothing more than autonomy.83 Some PJAK leaders do, however, retain a desire to create a greater Kurdistan.84 Prior to the onset of the Syrian crisis in 2011, Syrian Kurdish parties sought only greater recognition and political representation for Syria’s Kurds, not autonomy or independence.85 This level of ambition has increased only slightly as a result of the crisis.86

Therefore, van Bruinessen’s observation appears valid. In which case the creation of a Kurdish state could be thwarted by a lack of ambition to create one. However, there are a number of explanations for these relatively modest ambitions. The first is that most regional and international powers are opposed to Kurdish independence.87 Kurdish political groups might, therefore, disguise their true aims in order to appease their potential allies or deter a state response. For example, the president of the KRG, Barzani, has refrained from declaring independence from Iraq, knowing that it would threaten his party’s ties with Ankara and would not be supported by the US.88 A second explanation is pragmatism.89 This relates to Kurdish experiences of failed uprisings, repression and the above mentioned unwillingness of states to support the establishment of a Kurdish state. A final explanation is how opportunities within each state have shaped the demands of the Kurdish movements. David Romano has observed that Kurdish movements have been quick to exploit any political opportunities presented to them.90 This explains how Syria’s Kurds have been quick to mobilise as a result of the Syrian crisis, and how Iraq’s Kurds came to be in such a strong position today as a result of the series of openings presented to them since the 1990s. Therefore, should the right opportunity present itself, the Kurds could adapt quickly to exploit it.

**Summary**

For most Kurds, linguistic and religious differences no longer matter. Furthermore, the fact that many do not appear to seek independence now does not mean that they would not pursue it, should the right conditions emerge. However, two major internal impediments to the creation of an independent Kurdish state remain. The first and most significant is tribalism and the tribal behavior of political groups, which causes disunity among the Kurds. Until these groups can subordinate their own self-interests to the broader Kurdish nationalist cause, the Kurds ability to exploit any opportunities for independence will be severely impaired. Moreover, as the examples demonstrated, states can readily exploit divisions between Kurdish political parties in order to contain the Kurds and keep them divided. Second is the territory associated with the Kurds. Both the topography and the international borders within it have broken the coherence of the Kurdish nation. It also contains valuable natural resources which states will not wish to give up. Most significantly, it is the sovereign territory of Turkey, Iraq, Iran and Syria, all of whom see the preservation of their territorial integrity as vital to their existence. This will be explored further in the following chapters.

**Chapter Two - The international system**

Chapter two will examine the influence of the international system and external actors on Kurdish nationalist goals. It will show that the Kurds have historically ranked lower than the other interests of the major powers within the region. It will then explain why support is necessary to overcome the inherent resistance of the international system towards the creation of any new state, and why it is unlikely to be forthcoming.
The major powers

Britain, the Soviet Union and the US have all allied themselves with the Kurds at different stages during the twentieth century in order to further their interests within the region.

Britain sought influence among the Kurds while consolidating its position in Mesopotamia after the defeat of the Ottomans in 1918.91 It supported the provision for Kurdish autonomy, leading to independence one year later that was included in the Treaty of Sevres, signed in 1920 between the Ottomans and the Allies.92 However, no mention was made of the Kurds in its replacement, the Treaty of Lausanne, signed in 1923 with Mustafa Kemal.93 It is claimed that Britain dropped its commitment to the Kurds during the Lausanne negotiations in return for Turkey becoming a member of the League of Nations, and thus strengthening the isolation of Bolshevik Russia.94 In fact Britain had most likely already abandoned its commitment to Kurdish independence in 1922, when deciding to include the vilayet of Mosul, which contained the majority of Iraq's Kurds, in Iraq.95 To Britain, “Iraq…was not viable politically, militarily or economically without [it]”.96 Britain therefore strongly resisted Turkey’s claims to the vilayet of Mosul at the Lausanne negotiations.97 In the absence of a solution at Lausanne, a League of Nations commission eventually ruled in Iraq's favour. Britain abandoned the Kurds in 1932 once Iraq was admitted to the League of Nations.98 Ultimately, the appeasement of Kemalist Turkey, the rise of Bolshevik Russia, the viability of Iraq, and Britain's reputation in achieving the latter were all more important to Britain than Kurdish independence.99

While occupying northern Iran during the Second World War, the Soviet Union provided the support necessary for Iran's Kurds to establish the independent Republic of Mahabad. David McDowell judges that the Soviets wished to use Kurdistan and Azerbaijan as bargaining chips to extract oil concessions from Tehran. With Soviet backing, Qazi Muhammed, the KDPI leader, declared independence on 15 December 1945. Bound by treaty to leave Iran six months after the cessation of hostilities between the Axis and Allied powers, the Soviets departed in May 1946, but only after agreeing the formation of the Soviet-Iranian oil company with Tehran. Without Soviet support, Mahabad fell back under Tehran's control on 14 December 1946.100

“On three occasions during the first 50 years after 1945, Washington used the Kurds to further US interests in the Middle East and then cynically abandoned them to their own fate”.101 The two most infamous episodes occurred during the 1970s and 1990s. In 1972 the US, in conjunction with the Shah of Iran and Israel, provided support to the KDP in order to destabilise Saddam Hussein, who had just signed a treaty with the Soviet Union.102 Neither the US or Iran had any intention of allowing the Kurds to gain autonomy.103 Moreover, once Iran signed the Algiers Accord with Iraq in 1975, the US abandoned the Kurds, ignoring their pleas for support as they faced the consequences.104 Similarly, in the wake of Operation Desert Storm in 1991, the Kurds rose up in Northern Iraq based on comments emanating from the US that encouraged the Iraqi people to topple Saddam Hussein. When Iraqi forces retook the initiative, the US initially refused to support the Kurds for fear of creating a vacuum in Iraq that they were unprepared to fill.105

In all of these examples, Kurdish interests ranked below the other interests of Britain, the Soviet Union and the US. Furthermore, the Kurds position appeared to worsen in the wake of each episode. As Munir Morad put it, “[the] Kurds have failed to realise…that the superpowers are unlikely to act humanely against their best interests in either Baghdad or Ankara (or Tehran for that matter)”.106

Self-determination and territorial integrity

There are two factors at the international level that stand between the Kurds and statehood. The first is the conflict between territorial integrity and self-determination. The second is the requirement for international support. Both will be considered in this section, which will draw upon the conclusions made above when determining the likelihood of the Kurds achieving the latter.
The principle of self-determination is firmly enshrined within the UN charter and a number of UN treaties and declarations. Yet fundamental issues arise when trying to apply it. The first issues relate to the ‘peoples’ to which the principle should apply. In cases of national self-determination it is presumed that the ‘peoples’ are the nation itself. The Kurds offer an example of how challenging that is. Clearly defining who is and who is not a Kurd is complex, as the beginning of chapter one made clear. It is therefore difficult to determine who exactly is entitled to a say in the future of the nation in question, and where the territorial boundaries of that nation lie. A further problem linked to territory is the issue of how to deal with the other minorities contained within it. For the Kurds this would include Turcoman, Assyrians, Armenians, Yazidis, Turks, Arabs and Persians. In theory, these groups should be extended the same rights as the nation that surrounds them.

If a nation seeks statehood when exercising its right of self-determination, secession becomes the next major issue. The international system is disinclined to support secession for two reasons. The first is instability. This is caused by both the violent means that nationalists might pursue secession, and also by the entry of an endless amount of states into the international system should every eligible group pursue statehood. The second relates to the tension between territorial integrity and self-determination. It is now widely accepted that the former has precedence over the latter. Mikulas Fabry’s comprehensive study into the establishment of new states reinforces this point. He concluded that “[state] recognition practice of the post-Cold War period has preserved the supremacy of the principle of territorial integrity…over self-determination”. For all the reasons above, many scholars conclude that something short of secession is required when seeking to fulfil a nation’s claim to self-determination.

Studies and recent examples suggest that international support is necessary to achieve independence for three reasons. The first is to reverse the above mentioned paradigm that territorial integrity has primacy over self-determination. Kosovo, whose independence is opposed by a number of states because of that paradigm, is illustrative of this. A second situation where substantial external support would be necessary is if the Kurds had to fight militarily for their independence. Finally, a Kurdish state would have to be recognised by other sovereign states before the Kurds could enjoy the full benefits of independence, such as the ability to join institutions, sign treaties, undertake normal diplomatic and economic relations, and be protected by international law from acts such as external intervention.

Kurdish leaders recognise the requirement for external support for the reasons mentioned above, and because of the isolated position of Kurdistan. During an interview with Time magazine in 2012, the KRG President, Nechirvan Barzani said:

“How we can – I mean an independent Kurdistan – first of all we have to convince at least one country around us. Without convincing them, we cannot do this. Being land locked we have to have a partner, a regional power to be convinced and internationally, a big power to be convinced.”

That Kurdish groups have repeatedly sought external alliances, despite the many setbacks outlined at the start of this chapter, which is evidence of the enduring validity of this statement. Kurdish leaders and scholars therefore agree that without external support, it is highly unlikely that the Kurds could achieve independence and survive as a state.

However, it appears that external support for a Kurdish state is unlikely to be forthcoming. In the historical examples above it was shown that Kurdish interests were always subordinate to the other interests of the major powers. This remains the case today. Both the US and UK have explicitly stated a commitment to maintain the territorial integrity of Iraq. Russia has explicitly stated that it wants “Syria to remain a sovereign [and] territorially integral state”. There is no reason to believe that the US, UK or Russia would want anything different for the other states not mentioned. Therefore, in the current climate, it is difficult to foresee that the major powers would prioritise the creation of an independent Kurdish state over their commitment to maintain the status quo within the region.

Thus the argument so far suggests that the international system is not likely to support Kurdish independence. Gareth Stansfield, however, takes a more positive view. He sees potential parallels
between the Kurdistan region in Iraq and the trajectory that Kosovo took towards achieving its ‘independence’ in 2008.\textsuperscript{119} A brief examination of the ‘unique case’ of Kosovo, might therefore disprove the notion that the international system is not likely to support the creation of a Kurdish state. James Ker-Lindsay believes that Kosovo’s case can be reduced to three key arguments.\textsuperscript{120} The first is that Kosovo had a right to independence because of its special status as an autonomous province within Serbia prior to Yugoslavia’s disintegration.\textsuperscript{121} By this logic, the KRG could claim that it also deserves independence based on its autonomous status within Iraq. The second argument is based upon the massive human rights violations committed by Serbian forces in Kosovo.\textsuperscript{122} Again, the Kurds suffered comparable human rights violations under Saddam Hussein. The third argument was that returning Kosovo to Serbia after it had been under UN administration for so long would be impossible.\textsuperscript{123} It could perhaps be argued that Baghdad had remained under US supervision for so long, while the Kurds remained largely outside of it, although this is weak in comparison to the other arguments. The end result is that the Kurds could justifiably claim a right to independence by applying similar logic to that which was applied to Kosovo. However, one fundamental difference remains. Kosovo had the backing of over ninety countries, including the US, UK and France, and its independence is still not internationally recognised.\textsuperscript{124} It would take something significant to convince those same powers that the creation of a Kurdish state is more important than the preservation of the extant international borders within the region.

### Summary

The international system presents a substantial barrier between the Kurds and statehood for two reasons. First, Kurdish interests have always been subordinate to the other interests of the major powers in the region. Less a significant change in circumstances, such as a humanitarian disaster or genocide, it is unlikely that support for the creation of a Kurdish state would materialise. Second, recent experience shows that only in the most extraordinary circumstances would the international system prioritise a claim for self-determination above territorial integrity.

### Chapter Three – Turkey, Iraq, Iran and Syria

Two factors have been identified so far that would appear to prohibit the creation of an independent Kurdish state. These are disunity and a reluctance of the international system to prioritise claims of self-determination above the preservation of territorial integrity. It has also been evident that each state has resisted, and would continue to resist, the loss of any sovereign territory to a Kurdish state. This chapter will therefore identify what it is within each state that appears to impede progress towards an improved position for the Kurds, be that independence or otherwise. For each state, a brief historical overview will be followed by an analysis of what are deemed to be the main limiting factors.

#### Turkey

With approximately 15 million Kurds that account for about 20% of the total population, Turkey has the largest Kurdish population and the greatest part of the territory of Kurdistan of all four states.\textsuperscript{125} The construction of a Turkish nation-state that began under Mustafa Kemal Ataturk, during the 1920s, required the assimilation of the Kurds and other non-Turkish minorities into a single Turkish national identity.\textsuperscript{126} Assimilation measures included the establishment of Turkish as the sole official language, the introduction of a state controlled education system, the denial of access to government appointments by Kurdish officials, and the replacement of Kurdish names and place names with Turkish names.\textsuperscript{127} Further measures included forced deportation, detention without trial,\textsuperscript{128} executions,\textsuperscript{129} the destruction of Kurdish villages and the use of Turkish settlers to ‘Turkicize’ Kurdistan.\textsuperscript{130} These measures could rightly be perceived as an attempt to erase Kurdish identity altogether.\textsuperscript{131} From the 1980s, the PKK formed the main Kurdish opposition. Its tactics and maximalist claims for Kurdish independence elicited a strong securitised response from the Turkish state. A pattern emerged whereby Kurdish resistance to Turkish assimilation efforts led to a stronger state response, which in turn fuelled the strength of the Kurdish resistance.\textsuperscript{132}
From 1999 onwards, the prospect of Turkey’s accession to the EU, the capture of the PKK leader Ocalan, and the arrival of the AKP to power, marked a shift away from the militarised approach that had been adopted by the Turkish government until that point. In 2002, a harmonisation reform package, designed to meet the accession criteria for the EU, eased restrictions on the Kurdish language, Kurdish names, and broadcasting in Kurdish. The state of emergency in the last two remaining states was also lifted later that year. Viewing a political solution as more favourable than a military one, the AKP has overseen a number of events, and has introduced a number of reforms, that have directly or indirectly benefited the Kurds. These have included the first explicit reference to the Kurdish question by a Turkish Prime Minister in 2005; a ‘democratic opening’ in 2009 that sparked a public debate on the Kurds; and constitutional amendments that have removed Kemalist influence from the judiciary and the Army from Turkish politics. The AKP government also took the landmark step of engaging in talks with the PKK and Ocalan. These led to the announcement of a PKK ceasefire in March 2013, which initiated what is now referred to as the Turkish-Kurdish peace process.

Yet harsh measures towards the Kurds have continued parallel to these achievements. Military activity has continued in Kurdish areas. More significantly, in 2006 the Anti-Terror bill was amended to include the “broadest definition of terror the country has yet seen”. This has resulted in the detention of up to 5000 non-violent activists with alleged links to the PKK, among which include elected officials, journalists and academics. The AKP has also been accused of failing to address any of the real issues that affect the Kurds. Two explanations for this unsteady progress are Turkish nationalism and Kurdish nationalism. These appear to be the main factors that stand in the way of an improved situation for the Kurds within Turkey.

Turkish nationalism

Turkish nationalism is at the core of the nation-building project undertaken by Ataturk and his successors. Through education, a persistent state narrative, and an open media, that when reporting on the Kurds has taken its lead from the state, Turkish society and opinion has been shaped by the nationalist discourse since the birth of the Republic. This has been backed up by laws that prevent the emergence of alternative ideas. The most extreme Turkish nationalists have denied the existence of the Kurds, claiming that all Muslims of Anatolia were of Turkish origin. However, a more widespread belief is that all Kurds could become Turks. Indeed, it is argued that Kurds have the same rights as Turks, as long as they embraced Turkishness, which many have. Another belief shared by all Turkish nationalists is that Kurdish unrest is a threat to the territorial integrity of Turkey and is caused by external interference. This is widely known as ‘Sevres syndrome’, in reference to the promise of independence made to the Kurds in 1920.

Turkish nationalism has been particularly prevalent in state institutions, notably the military and judiciary. The military is credited with keeping the Kurdish issue as a security concern for so long, and the judiciary with preventing opportunities for a political solution. The latter has been achieved through the banning of Kurdish political parties by the constitutional court on the grounds of their alleged links to the PKK and the fact they are unconstitutional. This, combined with the above mentioned arrests of thousands of non-violent activists under the 2006 anti-terror law, has severely restricted opportunities for political dialogue. Incidentally, despite having the same effect, the 10% election threshold that has also kept Kurdish political parties out of parliament was not introduced with that purpose in mind. As self-appointed guardians of Kemalism and, until recently, through its highly influential role in state decision making via the National Security Council, the military has long viewed the Kurdish issue as a threat to the territorial integrity of Turkey and has remained a strong advocate of a security based solution.

Many agree that overcoming these deeply embedded beliefs will be exceptionally challenging. As Neophytos Loizides puts it: “what is…key in analysing Turkey is to understand that state ideology is entrenched in institutions including the military and judicial bodies that prevent reformulations of the national consensus even when key political figures realized the futility of state ideology”. This is evident in the political and public opposition to both the Turkish Prime Minister’s 2005 address and the
2009 Kurdish opening. Fear of “nationalist backlash” is also believed to be a reason why the AKP government has hesitated to implement reforms.

However, there are indicators that the influence of Turkish nationalism is weakening. As already mentioned, the AKP has ended Kemalist influence over the judiciary and removed the Army from Turkish politics. There was also reportedly no major public backlash to the government’s negotiations with the PKK, or to the democratisation package announced by the Turkish Prime Minister in September 2013. Furthermore, a survey reported by the International Crisis Group found that “only half the population or less is primarily aligned with ideologies that usually oppose reforms needed to address Kurdish grievances”. Judging by the AKP’s stuttering progress towards the Kurds, however, it appears that nationalism remains a powerful force in Turkish politics and society.

Kurdish nationalism

A significant factor that can inflame Turkish nationalist sentiment is Kurdish nationalism. Kurdish nationalists have frequently undermined public support for their cause by their actions, thus limiting the government’s freedom to make reforms. Examples include: PKK activity disrupting the Oslo process and the 2009 Kurdish opening; a BDP member of parliament causing “resentment and anger among the Turkish public” by praising a PKK suicide bomber that had caused the death of eight Turkish soldiers; and demands of “self-determination for Kurds through autonomy, federation or independence” made during a Kurdish political conference held in Diyarbakir in 2013, which provoked a significant Turkish nationalist response on social media.

Summary

The creation of a Kurdish state out of Turkish sovereign territory is an unrealistic proposition. The PKK and other Kurdish political groups have revised their goals to reflect that. Improved cultural and political rights are, however, achievable. Yet there appears to be two key factors that will impede progress in this direction. The first is Turkish nationalist sentiment. Changing the deeply embedded beliefs within the Turkish state and society will be no easy task. The government may identify reforms necessary to accommodate the Kurds, but will struggle to implement them if forced to adopt policies to win votes and appease Turkish nationalists. The second factor is the Kurds themselves. Provocative actions by Kurdish nationalists provide ammunition to Turkish nationalists and can also undermine public support for any reforms that the state may wish to introduce.

Iraq

Iraq’s Kurds number approximately three and a half to four million and, therefore, constitute about 15-20% of Iraq’s total population. Since Iraq’s formation its Kurds “have been in an almost constant state of revolt”. Four historical observations help place the current situation in context. First, Iraq’s Kurds have never forgotten the abandoned promises to protect their status and rights that were included in the 1922 Anglo-Iraqi statement, and by the League of Nations. Second, due to the artificial nature of Iraq, successive Baghdad governments have worked hard to maintain the Kurdish region under its control because it is vital for Iraq’s security and economy, and its separation could precipitate a further break-up of the country along sectarian lines. Third, tribalism has plagued Iraqi Kurdish politics and is exacerbated by external interference. Fourth, Iraq’s Kurds have been subjected to some extreme periods of violence by the Iraqi state, most notably during the Anfal campaign of 1988. Yet overall, Kurdish ethnic identity in Iraq has not been attacked to the same extent as it has been in Turkey. The invasion of Kuwait in 1990 was a seminal moment in Iraqi Kurdish history. The failed uprisings that followed led to the imposition of a ‘safe haven’ over Kurdistan. The autonomy enjoyed by Iraq’s Kurds today began in the space that this created. This is despite the opportunity almost being lost as a result of a civil war between the KDP and PUK between 1994 and 1998. Having experienced self-government since 1992, the Kurds were in a stronger position than their Arab counterparts when it came to forging
the Iraqi constitution in 2005. The “extremely decentralized federal system” that emerged now places the KRG in a very strong position.

The current situation in Iraq has been a cause for optimism among Kurdish commentators. One has gone as far as stating that “it would seem odd not to acknowledge [a future Republic of Kurdistan] as a distinct possibility”. However, it is judged that there are three factors that may dictate a different outcome for Iraq’s Kurds. These are internal divisions, the relationship between Baghdad and Erbil, and the influence of external actors.

Internal divisions

Gareth Stansfield stated in 2006 that “the greatest threat to the continued prosperity of the Kurdistan region in Iraq is the seemingly irreconcilable differences between the leaderships of the KDP and PUK”. In recent years, the rivalry between the two parties has been kept at bay because of the division of power achieved by the appointment of Talabani, the PUK leader, as the President of Iraq, and Barzani, the KDP leader, as the President of the KRG. Moreover, both parties have been notionally united within the construct of the KRG. However, divisions still exist, and conflict would not be difficult to foresee given recent shifts in the balance of power. Each party still maintains its own militia, and has separate relations with its neighbours; the KDP, based in the north, with Turkey, and the PUK, based in the west, with Iran. The PUK is losing political power, having come third behind the Gorran party during the September 2013 regional elections, and with its leader, Talabani, in poor health. It would not be implausible, therefore, to predict more internal conflict in a situation where rival parties based upon tribal allegiances control their own militias and maintain separate alliances with different regional rivals, particularly when one is losing political power.

A division also exists between Kurdish society and its political parties. There are two aspects to this. First is dissatisfaction with the rule of the KDP and PUK. Many view them as undemocratic, nepotistic and corrupt, which is in part the basis on which the Gorran party achieved its recent success. Second is the belief that the two parties fail to meet the expectations of the electorate. An unofficial poll held alongside the 2005 election found an overwhelming number of Iraqis sought independence. However, the mainstream parties, in public at least, remain committed to preserving the unity of Iraq.

Baghdad-Erbil relations

Disputes between Erbil and Baghdad, the most prominent of which are over oil and territory, threaten to undermine the Kurds position within Iraq. They have also revealed the extent to which the KRG is dependent on Baghdad for its economic wellbeing.

The KRG does not yet have the economic means to go it alone. It is currently allocated 17% of Iraq’s budget, most of which comes from Iraq’s southern oilfields. Until the KRG can match that allocation, it makes no sense to separate from Baghdad. Oil appears to offer the answer. However, the KRG’s ability to extract and export oil is contested by Baghdad. Regardless, the KRG has signed contracts with major companies and oil has been exported to Turkey. Baghdad has responded by using its leverage over the oil companies and by restricting access to Iraqi pipelines. With support from the US, it also persuaded Turkey to wait until the dispute is resolved before facilitating the export of Kurdish oil. Of greater impact, however, is the withholding of the KRG’s budget allocation. This has hurt the Kurdish region and has highlighted the KRG’s dependence on Baghdad, thus demonstrating its inability to prosper alone.

The disputes over territory within the provinces of Kirkuk, Ninevah and Diyala have many dimensions: territorial claims, historical injustices, oil and constitutional disagreements. Most significantly, it is these disputes that have brought the Kurdish Peshmerga and Iraqi National Army perilously close to conflict. The KRG is in a difficult position. It cannot be seen by Iraq’s neighbours and the US to be starting a conflict that might lead to the break-up of Iraq. The oil dispute has also demonstrated its economic dependence on Baghdad. Therefore, the KRG needs to find a way to resolve the disputes in
order to ensure the economic welfare of its people and to appease the international audience, or it risks jeopardising the strong position it currently enjoys. However, this will inevitably mean compromise on issues of great importance to the Kurdish population who, as mentioned above, are increasingly disillusioned with their leaders.

External actors

The traditional belief is that Iraq’s neighbours and the US would oppose an independent Iraqi Kurdistan in the interest of preserving stability, and to prevent it from becoming an inspiration for Kurds elsewhere within the region. This still appears to be the position of the US and Iran. Turkey, on the other hand, is less clear. Some question whether Turkey is coming to terms with the prospects of an independent Kurdish region. Reasons include Ankara’s deepening political and economic ties with the KRG; an alignment of interests, most notably energy; and, an improvement in Turkey’s approach to its own Kurdish population. The counter view is that Turkey would still resist the break-up of Iraq. The traditional fears of instability and the spread of Kurdish nationalist sentiment remain. Furthermore, Turkey has sizeable economic interests within Iraq outside of the Kurdish areas. Whatever the case, Turkey’s support has undoubtedly strengthened the position of the KRG in recent years, and it would be hard to imagine the KRG prospering without it in future. Yet Turkey’s support is still unlikely to extend to a position that would support independence. Some, therefore, believe that the only way support for an independent Kurdistan would materialise would be if Iraq broke apart, despite Kurdish efforts to prevent that from happening.

Summary

The Kurdish region in Iraq is in a very strong position relative to its Kurdish neighbours, and there appears to be very little that Iraq’s Kurds could do to better it. They are very unlikely to get the support necessary to bring about independence alone. Moreover, the oil dispute has shown that independence from Baghdad might not be in the Kurds best interest. Perhaps the greater challenge now would be to protect the gains made during the past decade. Key to this will be maintaining the support of Turkey, resolving the disputes with Baghdad and preventing internal divisions from getting worse. None of these factors are assured, and a failure in any one of them could result in a set-back for Iraq’s Kurds.

Iran

With between seven and nine million Kurds, Iran has the second largest Kurdish population behind Turkey. However, at only 12-15%, the Kurds make up a smaller proportion of the total population than in Turkey and Iraq. It is also important to note that the Kurds are one of many ethnic minority groups within Iran, and that Persians account for less than half of the total population. Nevertheless, Persian has been the dominant ethnic identity around which the state of Iran has been built. Beginning in the 1920s, Reza Shah, like Ataturk, sought to ‘modernise’ Iran. To the Kurds this meant the imposition of the Persian language, the settling of tribes, and the exclusion of Kurds from government appointments, among others. To differing degrees, successive governments have maintained similar policies. Moreover, approximately half of Iranian Kurds that are Sunni have suffered further discrimination on the grounds of their religion. However, it is generally believed that, at least culturally, Kurds in Iran have been afforded more space than Kurds in Turkey, Iraq and Syria. One suggested reason for this is the cultural similarities between Kurds and Persians, which include celebrations such as Newroz, and the relationship of the Kurdish language to Farsi.

A combination of the repressive policies of the state, and exposure to nationalist ideas, led to the emergence of Kurdish nationalism in Iran. Its crowning achievement was the establishment of the Republic of Mahabad in 1946. There have been a number of other uprisings, particularly during times when the Iranian state has been at its weakest. However, on the whole, Iranian Kurds have been unable to mount a serious or sustained challenge to the Iranian government. A variety of reasons have been given for this, including the proportionally small size of the Kurdish population and the cultural factors mentioned above. The two most significant factors, however, have been the strength of the state
response, and the fractured nature of the Iranian Kurds. It is proposed that these remain the key impediments to the Kurdish nationalist cause in Iran today.

Internal divisions

Tribal divisions were a major factor in the failure of the two major Iranian Kurdish uprisings of the early twentieth century, the Simko rebellion of 1925, and the Republic of Mahabad in 1946. Since the mid-1980s, fighting within and between political parties has continued to blight Kurdish progress. At its worse, this took the form of a war between the two main parties, the KDPI and Komala, in 1985. Splits within the KDPI in 1988 further complicated the situation. Even within the past decade, further splits have occurred within the KDPI and Komala.

The situation has been further exacerbated by the relationship between Iranian Kurdish nationalist groups and their counterparts in Iraq. During the late 1960s and 1970s, and again from the early 1990s until present, the KDPI and Komala have been prevented by the Iraqi KDP, then KRG, from fighting the Iranian government, because of the relationship the Iraqi groups had with Tehran. In the 1970s, the leader of the KDPI, Abdul Rahman Ghassemlou, wrote that "in the Kurds struggle against the Shah's regime, the democratic forces of Iran were more reliable and significant allies than even our fellow Kurds of Iraq or Turkey".

Since 2004, the PKK affiliated PJAK has added an additional dimension. Its choice of armed struggle over politics, at a time when all other Iranian Kurdish political groups had decided otherwise, has led to it being kept at a distance. In short, Kurdish Iranian political parties are exceptionally fractured and are, therefore, ill-disposed to present a united front against the Iranian government. Their position is further weakened by the methods employed against them by the state.

The Iranian State

The Iranian government has virtually neutralised the Kurdish nationalist movement as a military and political force within its borders. The former has been achieved by the use of overwhelming military force to counter any uprisings. For example, 250,000 Iranian troops, supported by air power, armour and artillery, were used to counter the Kurdish uprising attempted at the start of the Iran-Iraq War. A similar strength garrison was maintained within the Kurdish region during the 1990s. Similarly, an offensive by the Iranian military reportedly prompted the PJAK to declare a ceasefire in 2011, although it is unclear whether it was the offensive that prompted the ceasefire or Iranian collaboration with the PKK. Whatever the case, most, if not all Kurdish nationalist groups within Iran have realised that they do not have the strength to confront the Iranian military.

A variety of means have also been used to push Kurdish nationalists out of the political space. One of the most damaging has been the assassination of Kurdish leaders and officials. The KDPI was particularly affected by the assassination of their leader, Ghassemlou, in 1989 in Vienna, followed by his successor, Sadiq Sharafkandi in Berlin in 1992. Similarly, hundreds of other Iranian Kurds were assassinated by Iranian government operatives in the Kurdish region of Iraq during the 1990s. Further measures have included the prevention of Sunni Kurds from occupying government appointments, the banning of Kurdish political parties, and the use of executions on political prisoners and, disproportionately, against minority groups in general. These measures have pushed the Kurdish political movement underground, thus dislocating parties from those they seek to mobilise. They also serve to further shatter the already fragmented Kurdish political scene.

Summary

It is difficult to predict whether the two main impediments in Iran, the state and disunity among Kurdish political parties, are likely to improve. Iranian Kurds did begin to experience greater cultural rights under the reformists, Rafsanjani and Khatami, but progress was ultimately thwarted by hard-line conservatives. There was also early optimism that things would improve under the new Prime
Minister, Hassan Rouhani. However, so far he has, reportedly, failed to deliver his promise to improve minority rights, and executions have continued to rise since his election in June 2013. Moreover, the main Kurdish political parties still cannot operate within Iran, which therefore makes a political solution difficult to negotiate. Nothing therefore indicates that the state is willing to alter its position towards the Kurds. A political solution is further hindered by disunity among Kurdish political groups. Progress is being made, particularly within and between the KDPI and Komala, who sought to unite prior to the 2013 attempt to convene a pan-Kurdish conference. However, it is too early to tell whether this can last. Furthermore, both parties remain distanced from PJAK.

Syria

The Kurds in Syria number about one million and make up approximately 10% of the total population. They have also been subjected to discrimination and oppression, but not to the same extent as in Turkey, Iraq and Iran. The greatest period of repression occurred after the arrival of Arab nationalism from the mid-1950s onwards. The two most notorious acts introduced around this time were the removal of citizenship from as many as 150,000 Syrian Kurds in 1962, and the attempt to create an “Arab Cordon” along the Syria-Turkey border in 1973. However, on the whole, the response by the Syrian Kurdish nationalist movement has been weak and fragmented. That Syria’s Kurds are geographically divided between three unconnected regions and other isolated pockets must also complicate its ability to mobilise the Kurdish population. It is also believed that the presence of the PKK in Syria between 1980 and 1998 further curbed Kurdish nationalist activity against the Syrian government. The Syrian government had allowed the PKK to operate from Syria in its dispute over Turkey’s use of Euphrates River water and over the loss of the Sanjak of Alexandretta.

The Arab Spring in 2011 presented a number of opportunities for Syria’s Kurds. Initially, Assad sought to persuade the Kurds against joining the uprising by granting citizenship to those affected by the 1962 census. In general, however, the Kurds have neither aligned themselves with the Assad regime nor the ‘opposition’, both of whom are viewed as unsupportive of Kurdish aspirations. Kurdish political parties have instead sought to pursue their own agendas, which are many and varied. Kurds have been in control of Kurdish areas since the departure of Syrian government forces in mid-2012. The PYD, which is linked to the PKK and is widely believed to have at least the tacit support of the Assad regime, has emerged as the dominant group. This is of concern to Turkey, and is a source of tension between the PYD and the other Syrian Kurdish groups. Three key factors that will affect the future of the Kurds in Syria therefore emerge out of this situation. These are internal disunity, external actors, and the attitude towards the Kurds of whatever future Syrian government emerges out of the crisis.

Internal divisions

Significant divisions exist between the PYD and the Kurdish National Council (KNC), and also within the KNC. The KNC comprises 12-15 smaller parties that were brought together to counter the dominance of the PYD. It is far from united, with each party having a narrow support base, consisting of familial and territorial links, and different ideologies. It also has very little influence over the Syrian Kurdish population and no military strength. The KRG, under Barzani, has sought to address the latter by training KNC fighters in Iraq. However, the military arm of the PYD, the People’s Defence Corp (YPG) has prevented them from entering Syria. Therefore, the KNC’s only strength is the legitimacy it receives from the support provided by the KRG and Turkey.

The PYD is strongly believed to be organisationally, ideologically and militarily aligned with the PKK, but seeks to distance itself from that link. It is also the most well organised and widely supported Kurdish group within Syria. That, and its particularly effective military arm, the YPG, is the reason why it has proved so able to govern. These strengths, however, are negated by its links to the PKK, and its suspected alliance with the Assad regime. The latter is also disputed. There are examples of PYD forces fighting the regime. However, there are also examples of state complicity with the PYD. Regardless of the truth, the simple belief that the PYD is linked to the regime and the PKK is enough to
The KNC does not have the cohesion, strength or support base to effectively govern Syria’s Kurdish regions, but it does have the support of Turkey, the KRG and some Western powers. The PYD is well equipped to govern in Syria, but is isolated because of its association with the PKK and Assad regime. Disunity within the KNC, and the rivalry between the KNC and PYD, will therefore impede those groups ability to negotiate on behalf of Syria’s Kurds with any future Syrian government.

External actors

External actors, who seek to manipulate the outcome in Syria to their advantage, exacerbate the divisions mentioned above. Turkey is the most influential in this regard. It views the PYD and PKK as one and the same, so is opposed to the prospect of a PYD controlled Kurdish region on its border. It has attempted to influence events in the Kurdish regions by initially threatening the use of force, and then by using its influence over the Syrian opposition, as well as its alliance with the KRG, to control Kurdish aspirations. Turkey has also been suspected of supporting attacks by opposition groups, including Islamist groups, against the PYD, which it denies. Furthermore, it has sought to physically isolate the PYD by constructing a ditch along the international border that surrounds the Kurdish regions in conjunction with Iraq. There are, however, signs that Turkey is coming to terms with the PYD’s dominance. Turkish officials have opened dialogue with the PYD. Furthermore, the movement of humanitarian aid and Turkish military convoys has been coordinated between Turkey and the PYD. Nevertheless, it still appears that Turkey continues to prefer the KNC. Moreover, by talking to both sides, Turkey can keep the Kurds divided and thus undermine their attempts to gain autonomy in the long run. Of course, Turkey’s ability to cooperate with the PYD will also remain dependent upon the progress it can make towards its own Kurdish population and the PKK. The KRG has remained largely in step with Turkey over its backing of the KNC over the PYD. However, in recognition of his own experiences of the damage that internal rivalries can cause, and to Ankara’s discomfort, Barzani has sought, albeit unsuccessfully, to unite the PYD and KNC.

Summary

The future of the Kurds in Syria will depend on three factors. The first is the position that any future Syrian government chooses to take towards the Kurds; a factor that is incredibly difficult to predict. The second is Turkey. Turkey’s stance towards a future Kurdish entity in Syria will depend on its relationship with any future Syrian government and its relationship with its own Kurds. It will likely do its utmost to prevent the future use of Syria as a base for Kurdish nationalists. Turkey might, however, tolerate a Kurdish region on its border that is not pro-PKK, as it has with Iraq. The final factor is Kurdish unity. There appears to be a genuine risk that divisions within the KNC and between the PYD and KNC could undermine attempts to negotiate a future for Syria’s Kurds with the next Syrian government.

Turkey, Iraq, Iran and Syria have been, and remain, formidable opponents of Kurdish independence. When strong, as in Turkey and Iran, they also continue to resist progress towards Kurdish aspirations short of independence. In Turkey, the main impediment appears to be Turkish nationalism, which is easily aggravated by Kurdish nationalism. In Iran, a strong state that similarly views concessions towards the Kurds as a threat to state unity is also a major impediment, as is the fractured nature of Kurdish politics. The Kurdish region in Iraq is in by far the best position and is the most likely candidate to achieve independence under the right circumstances. However, Kurdish disunity and confrontation between Baghdad and Erbil threatens to undermine that. Turkey might also become an impediment to Kurdish aspirations within Iraq should it alter its position towards the KRG. In Syria, Turkey and disunity are also potential impediments to Kurdish aspirations, alongside the largely unknown factor of the future Syrian government. Three common impediments to both Kurdish independence and any aspirations short of
that have therefore emerged. These are disunity, Turkey, and state governments that are driven by ideologies that are opposed to any accommodation towards the Kurds.

**Conclusion**

This paper set out to identify the impediments to the creation of an independent Kurdish state. An analysis of the Kurdish nation and the international system unearthed two. The first was Kurdish disunity. The second was the strong inclination of the international system to protect the territorial integrity of Turkey, Iran, Iraq and Syria. The focus of the study then switched to analysing the position of the Kurds within each state, in order to identify the specific impediments to both Kurdish independence and, more realistically, Kurdish aspirations short of independence. Some common themes emerged from this. Kurdish disunity was again identified as a factor, as was the role of external actors. Turkey in particular was found to be playing a key role in Iraq and Syria. A final impediment was the inability of states to make reforms that could satisfy Kurdish aspirations.

The tendency of the international system to uphold the principle of territorial integrity over claims of self-determination is a considerable obstacle to Kurdish independence. In the absence of a profound change in circumstances, it is highly unlikely that the Kurds would gain the support necessary to reverse that precedent. The Kurds have historically ranked below the other interests of the major powers within the region, and there is nothing to suggest that that has changed. Most, if not all, states of any significance appear committed to preserving the territorial integrity of Turkey, Iraq, Iran and Syria. The resources and other minority groups contained within the greater Kurdistan region are other factors that strengthen the resistance to Kurdish independence.

Tribal and political disunity has been a recurring theme throughout this paper. It is both an impediment to Kurdish independence, and a factor that could undermine the Kurd’s ability to improve their position within each state. At the pan-Kurdish level, the greatest rift lies between the PKK in Turkey and the KDP in Iraq. Its effect reverberates in Syria between the PYD and KNC, and in Iran between the PJAK and the other Kurdish groups. It was also a major factor behind the failure to convene a pan-Kurdish conference in 2013. At the state level, the KDP-PUK rivalry remains a significant threat to Kurdish unity in Iraq, and the many non-PKK aligned parties in Iran and Syria remain notoriously fractured. Disunity is not as apparent in Turkey where the PKK and its affiliated organs dominate the Kurdish movement. However, that in itself is an impediment because more moderate voices are either pushed to the sidelines or incorrectly grouped under the PKK umbrella. The effects of disunity are threefold. It weakens the Kurds position vis-à-vis the state. Resource that should be used to better the Kurds position is instead directed towards internal fighting. Finally, it is a weakness that state governments can exploit to contain the Kurds at home or create instability in neighboring states.

The Turkish government seems to have considerable influence over the future of the Kurds within Iraq and Syria and, of course, over its own Kurdish population. Turkey’s relationship with the KRG appears key to the future prosperity of the Kurdistan region of Iraq, particularly while the KRG’s relationship with Baghdad is so fraught. Turkey is also attempting to influence events in Syria. There is also much that the Turkish government can do to keep the Turkish-Kurdish peace process on track. It is largely the AKP that has kept things moving, albeit slowly, in a direction that is favourable to the Kurds. Kurdish hopes, therefore, seem dependent upon the AKP remaining in power, or future Turkish governments adopting a similar or better position to it. This is not guaranteed, especially when considering that a lot of what has happened in recent years would have been almost unthinkable within Turkey prior to the late 1990s.

A final factor is the ability of state governments to make the reforms necessary to meet Kurdish aspirations short of independence. This mostly applies to Turkey and Iran, but it may also apply to a future Syrian government. Some generic reforms include lifting restrictions on the Kurdish language, allowing Kurdish political parties to operate legally, allowing Kurds access to the full range of government appointments, and ending the use of extraordinary security measures against non-violent political activists. It might also include the introduction of some form of autonomy for the Kurds. However, the ability of states to introduce reforms of this nature are resisted by strong internal forces; for
example, Turkish nationalism in Turkey and hard-line conservatives in Iran. Moreover, insensitive acts by Kurdish nationalists only serve to strengthen those forces of resistance.

Events in Iraq, Syria and Turkey have resulted in an improved position for the Kurds within those states. However, progress is slow and by no means assured, and the impediments identified in this paper suggest that an independent Kurdish state still remains unlikely. Only an extraordinary set of circumstances could change that. This does not necessarily mean that Kurdish nationalism is set to remain a threat to stability within the Middle East. The fulfilment of Kurdish aspirations short of independence could mitigate that prospect. Yet impediments to even modest concessions towards the Kurds remain within each state, as this paper has identified. Thus the Kurds will likely remain a source of instability in the Middle East as long as their aspirations are not being met.

Endnotes

2 MacDowell, *A Modern History*, xi.
7 Ibid., 452-453.
8 Ibid., 453.
10 van Bruinessen, “Kurdish Paths to Nation,” 23.
12 van Bruinessen, “Kurdish Paths to Nation,” 23.
13 James Ker-Lindsay, “Preventing the Emergence of Self-Determination as a Norm of Secession: An Assessment of the Kosovo ‘Unique Case’ Argument,” *Europe Asia Studies* 65, no. 5 (2013): 839.
15 Ibid., 13.
16 Ibid., 12.
20 van Bruinessen, “Kurdish Paths to Nation,” 27.
22 Ibid., 198; and van Bruinessen, “Kurdish Society,” 38-40.
24 van Bruinessen, “Kurdish Paths to Nation,” 25.
25 Ibid., 41.
26 Ibid., 36.
30 Ibid., 203 and 212-213.
33 van Bruinessen, “Kurdish Society,” 36.
38 For examples see MacDowell, A Modern History, 132, 169 and 385-391.
40 Stansfield, “Finding a Dangerous Equilibrium,” 266.
41 van Bruinessen, “Kurdish Society,” 42.
46 Ibid., 227.
47 Ibid., 229.
48 van Bruinessen, “Kurdish Paths to Nation,” 36.
49 Ibid., 36.
51 van Bruinessen, “Kurdish Paths to Nation,” 36.
52 Friej, “Alliance Patterns,” 24, 29-32.
53 Ibid., 31.
54 Ibid.
58 Park, “Turkey’s Multiple Kurdish Dilemmas,” 44.
63 Ibid.
67 Friej, “Alliance Patterns,” 22.
69 Friej, “Alliance Patterns,” 32.
71 Park, “Turkey’s Multiple Kurdish Dilemmas,” 44.
72 MacDowell, A Modern History 480; Park, “Turkey’s Multiple Kurdish Dilemmas,” 44.
74 Ibid.
75 ICG, “Turkey; The PKK and a Kurdish Settlement,” 29.
76 Updegraff, “Turkey Under the AKP,” 120.
82 Ibid., 270.
88 Park, “Turkey’s Multiple Kurdish Dilemmas,” 55.
91 See MacDowell, *A Modern History*, 115 and 117.
92 Ibid., 115 and 464. For criticism of the treaty see Nezan, “The Kurds,” 34-35.
96 MacDowell, *A Modern History*, 143.
97 For Turkey’s claims see Ali, “The Kurds,” 521-522.
98 MacDowell, *A Modern History*, 177.
100 MacDowell, *A Modern History*, 239-245.
102 Ibid., 77.
103 Ibid., 83; and Friej, “Alliance Patterns,” 25.
104 Little, “The United States,” 83.
105 Ibid., 91.
107 Article 1(2) and Article 55 of the UN Charter; Article 1 of the International Covenant on Civil and Political Rights, and the International Covenant on Economic, Social and Cultural Rights; UN Declaration on the Granting of Independence to Colonial Countries and Peoples, 1960; and UN General Assembly Resolution 2625 [XXV], 1970.
109 Ibid., 91 and 94.
110 Ker-Lindsay, “Preventing the Emergence of Self-Determination,” 839-841; Kirisci and Woodrow, *The Kurdish Question*, 54.
116 Newton-Small, “An Interview with Nechirvan Barzani”.
119 Stansfield, “The unravelling,” 282: Noting that Kosovo’s independence is still not recognised by a majority of the members of the UN.
120 Ker-Lindsay, “Preventing the Emergence of Self-Determination as a Norm,” 847.
121 Ibid.
122 Ibid., 848.
123 Ibid., 849.
124 Ibid., 838.
125 Gunter, “The Kurdish Question,” 197 and 198.
129 MacDowell, *A Modern History* 196.
130 Ibid., 208-209.
131 Ibid., 210.
133 Updegraff, “Turkey Under the AKP,” 123.
134 Park, Modern Turkey, 97.
137 Ensaroglu, “Turkey’s Kurdish Question,” 13-14.
138 Ibid., 15.
140 Ibid., 20.
143 Ibid., 19.
144 Henri Barkey and Graham Fuller, Turkey’s Kurdish Question (Oxford: Rowman and Littlefield, 1998), 121-122.
145 Park, Modern Turkey. 23-24.
147 Ibid., 137.
148 Ibid., 138.
150 MacDowell, A Modern History, 441.
153 Park, Modern Turkey, 60-61, 83.
156 Ensaroglu, “Turkey’s Kurdish Question,” 12; Park, “Turkey’s Multiple Kurdish Dilemmas,” 50.
157 ICG, “Crying ‘Wolf’“, i and 11.
158 Ibid., i; Ensaroglu, “Turkey’s Kurdish Question,” 13.
161 Ensaroglu, “Turkey’s Kurdish Question,” 13; Park, “Turkey’s Multiple Kurdish Dilemmas,” 50.
165 Ibid., 201.
166 See MacDowell, A Modern History, 151-180 and 287-363.
167 Romano, The Kurdish Nationalist Movement, 216.
168 Gunter, “The Kurdish Question,” 201
169 See MacDowell, A Modern History,357.
171 Romano, The Kurdish Nationalist Movement, 217.
175 Stansfield, “Finding a Dangerous Equilibrium,” 274.
181 Gunter, “Arab-Kurdish Relations,” 1624.
182 Gareth Stansfield, ‘From the ‘Kurdish Issue’ to the ‘Kurdistan Region of Iraq’,” in Iraq Ten Years on, eds. Claire Spencer et al. (London: Chatham House, Royal Institute of International Affairs, 2013), 25.
184 Ibid; Raad Alkadari, “Oil and the question of federalism in Iraq,” International Affairs 86, no. 6 (2010): 1319-1320.
186 Alkadari, “Oil and the question of federalism in Iraq,” 1322.


196 Idiz, “Turkey defers to Baghdad on oil from Kurdistan”.

197 Stansfield, “The unravelling,” 266.


199 Yildiz and Taysi, *The Kurds in Iran*, 3.

200 Ibid., 35.

201 Ibid., 50.

202 Ibid., 5.


204 Koohi-Kamali, “The Development of Nationalism in Iranian Kurdistan,” 179.

205 Ibid., 175.


207 Romano, *The Kurdish Nationalist Movement*, 244.


214 Ibid., 15-16, 21.

215 Ghassemlou, “Kurdistan in Iran,” 120.


222 Ibid.


228 Gunter, “The Kurdish Question,” 198.

229 MacDowell, *A Modern History*, 476.


232 Aylin Unver Noi, “The Arab Spring, its effects on the Kurds, and the approaches of Turkey, Iran, Syria and Iraq on the Kurdish Issues,” *Middle East Review of International Affairs* 16, no. 2 (2012): 17.


234 Ibid., 479-480.


236 Gunter, “The Kurdish Spring,” 447; Park, “Turkey’s Multiple Kurdish Dilemmas,” 44.


243 Ibid., 4.
244 Ibid., 25.
245 Ibid., 18.
251 Park, “Turkey’s Multiple Kurdish Dilemmas,” 44.
256 Park, “Turkey’s Multiple Kurdish Dilemmas,” 47.
260 Park, “Turkey’s Multiple Kurdish Dilemmas,” 47.
262 Zaman, “Syrian Kurd leader”.
264 Park, “Turkey’s Multiple Kurdish Dilemmas,” 57.
265 Ibid., 47.
China’s peaceful rise or otherwise:
Implications for security and prosperity in the Asia-Pacific

COL Chris Websdane, DSC, Australian Army

Introduction

The economic growth of China since the death of Mao Zedong in 1976 has been nothing short of phenomenal. Following systemic reform instituted by Deng Xiaoping in 1978, the Chinese economy has transformed into the second largest economy in the world in terms of Gross Domestic Product (GDP), with China now also the world’s largest trading nation. For three decades, according to Chinese data, growth was recorded in double-digit figures and, although slowing to 7.6 percent in 2013, China is predicted to surpass the United States as the largest economy in the world by 2015.

Backed by this long-term economic development, China’s military power has also been rapidly growing. The pace and scope of this modernisation, together with a lack of transparency over the military budget, and recent antagonistic actions over disputed territories in the South China and East China Seas, has generated concern with respect to how China might use these capabilities in the future. The possibility of a redefined Chinese strategic outlook has generated further uncertainty with respect to geo-political power relationships in the Asia-Pacific. What are China’s regional ambitions and what are the implications for United States strategic interests, both in terms of hegemonic influence in the Asia-Pacific, but also as the unipolar superpower in global affairs? The 2011 announcement of the return of US focus to Asia is a further dimension of the geopolitics in the region, as well as a potential source of friction and future conflict between the two powers. Adding to the complexity of the situation are serious internal social and political challenges confronting the Chinese Communist Party (CCP) that, unaddressed, could affect economic stability in China, with possible economic and security consequences for the region.

For a middle power with significant economic and security engagement in the Asia-Pacific, the continued “peaceful rise” of China is a matter of utmost strategic importance for Australia. The potential for conflict between Australia’s most important trading partner and its closest security ally is a scenario that would have a devastating impact on future Australian prosperity. This potentiality, and the magnitude of its consequences, has generated much debate with respect to Australia’s strategic approach to the region, and engagement with these two countries. Does Australia need to make a choice between the US and China, or can it use its influence with both countries to achieve a balance, that may facilitate longer-term regional and global stability?

The purpose of this paper is to identify the implications for security and prosperity in the Asia-Pacific as a result of the re-emergence of China as a great power. The essay examines the evolution of Chinese foreign policy in the region and the current and future sources of geo-political tension between China and the US, and China and the region. It identifies three possible scenarios for future Chinese participation in regional security arrangements, and then briefly considers the implications for Australia in balancing relations between China and the United States.

Emergence from isolation

Viewed through the perspective of history, the re-emergence of China to great power status in the 21st Century should not be as surprising or unexpected as much of the literature on the subject suggests. The rudiments of the Chinese Empire were beginning to take shape 3,000 years ago. A succession of Dynasties used military, cultural and economic means to achieve territorial expansion
and extend Chinese influence to all parts of the civilised world. During the early part of the Ming Dynasty, China was the hegemonic power of East Asia and maritime South East Asia; a power status achieved through aggressive foreign policy. Although imperial fortunes of the Ming waned, the Qing Empire reasserted China’s power in the 17th Century and for the next 150 years, dominated East Asia. Prior to the onset of the Second World War, China was still a powerful nation in its own right. China accounted for 2-3 percent of global trade and was ‘politically credible enough’ for her inclusion as a great power in the post-war international order. What makes this most recent rise so impressive, however, is the speed at which China has developed economic capacity, given the lengthy and debilitating period of international isolation under Mao.

The catalyst for change was the ascension of Deng Xiaoping to the position of paramount leader in 1978. Deng was the architect of a range of radical reform policies that established the springboard for China’s rapid and astonishing re-emergence to “great power” status. Deng had two objectives: to make China strong again and to preserve the primacy of the CCP. Deng realised the path to a strong China was through a strong economy, but to achieve this economic development, China needed access to foreign markets, technology and investment. Consequently, Deng instituted his policy of ‘opening up and reform’, and abandoned the Soviet-style planned economy, in favour of a ‘socialist market economy’.

Evolution of foreign policy

As significant as China’s economic rise has been, the re-engagement of China with the international system following the death of Mao has been equally meteoric. Under Deng’s reformist policies, the main objective of China’s foreign policy was to create conditions advantageous for economic development, in order to achieve his “Four Modernisations” strategy: modernisation of industry, agriculture, science and technology, and national defence. Fundamental to realising this objective was peaceful relations with the great powers and the existence of a peaceful and stable international environment. In contrast to Mao Zedong’s key assumption ‘that war was inevitable and China must be prepared against war’, Deng made the assumption that world war was unlikely to break out in the foreseeable future. Consequently, economic development became a core objective of the CCP over preparing for war, and economic modernisation the primary focus of China’s reform-era foreign policy.

To facilitate economic growth, Deng accelerated engagement with the international community. Between 1988 and 1994, Beijing established diplomatic relations with 18 countries. Similarly, in efforts to support a peaceful and stable environment, China actively engaged bilaterally to resolve territorial disputes with Kazakhstan, Kyrgyzstan, Laos, Russia, Tajikistan and Vietnam, in some cases, achieving outcomes not necessarily to China’s advantage. China also broadened engagement into multi-lateral forums, as a means to pursue trade and security interests, as well as to hedge against American regional influence and attempt to avoid US-led containment. China increased interest in, and commitment to, regional multi-lateral organisations, including the Asia-Pacific Economic Cooperation (APEC), the Association of Southeast Asian Nations Regional Forum (ARF) and the East Asia Summit (EAS). In 1996, as a counter-balance to US economic and security influence in the region, China formed the Shanghai Five with Russia, Kazakhstan, Kyrgyzstan and Tajikistan (changed to the Shanghai Cooperation Organisation in 2011 following the inclusion of Uzbekistan). At the same time, China commenced its bilateral “partnership diplomacy” approach, establishing partnership relationships with Russia (1996), the US (1997) and Japan (1998). This dramatic and comprehensive opening of China, in particular from 1992 after normalisation of international opinion following the Tiananmen incident, led to a significant increase in foreign trade and foreign direct investment, with profound socioeconomic benefits. In 2001, China entered the World Trade Organisation.
International perceptions

As Chinese interaction with, and re-integration into the international community increased and Chinese diplomats gained more experience, Chinese officials became more attentive to foreign perceptions of Chinese policy, in particular the views of some countries within Asia that saw China as a potential source of regional instability. China’s occupation of Mischief Reef in 1995, missile firings and military exercises in the Taiwan Straits, and bellicose language towards Taipei, generated serious alarm amongst regional neighbours such as Japan and South Korea. This apprehension, further clouded against the opaque meaning of Deng’s maxim “hide your strength, bide your time”, was counter-productive to economic development.

In an attempt to allay fears of the US and regional neighbours, Zheng Bijian termed the phrase “peaceful rise” to clarify China’s extraordinary economic and military advances of the previous decade. Through ‘the development path to a peaceful rise’, Zheng articulated Chinese intention to take advantage of the peaceful international environment to continue to modernise, but also to enable China’s development to contribute to sustaining international stability and economic prosperity.

In an attempt to further avoid misinterpretation of the term “rise” and to reassure the international community of China’s peaceful intentions, President Hu Jintao adopted the “peaceful development” narrative in 2004. Reinforcing “peaceful co-existence” and “peace and tranquillity”, this strategy reiterated the maintenance of a stable international and domestic environment as essential preconditions for continued Chinese development and long-term prosperity. Official narrative also made clear China’s intentions regarding regional or global leadership, in that (China would) ‘never seek hegemony, no matter how well-developed it becomes.’ The subtle change in narrative tacitly, if not begrudgingly, acknowledged China’s acceptance of the unipolar international system and continued US hegemony in Asia. It also signalled China’s role as a rising power in regional and world affairs.

National interest versus great power influence?

From 1978 until the early 2000’s, China’s approach to foreign policy was characterised by the adoption of the “opening and reform” course. Kitano described this approach to foreign policy as ‘internationalism in the context of emphasising domestic re-vitalisation’. In this regard, although China had opened up to the international system economically to achieve its modernisation objectives, and was acutely more aware of her diplomatic influence, in the political realm, China was considerably less amenable to international values and norms. This was particularly the case with their approach to democracy and human rights. Furthermore, tending to maximise its interests through minimal involvement, ‘Beijing sought many of the rights and privileges of a great power, without accepting most of the attendant obligations and responsibilities.’ In so far as engaging with the international system is concerned, China has sought opportunities, not responsibilities.

From 2006, China commenced redefining its national interests, adding national sovereignty and security to the economic considerations for maintaining advantageous conditions for economic development. During this time, the term “core interests” increasingly came to be used in bilateral language as a means to articulate those issues that China will not compromise on. By 2009, China’s three core interests were articulated as the preservation of the political system and state security, state sovereignty and territorial integrity, and continued stable economic and social development.

Coinciding with this redefinition of national interest has been an increasing assertiveness with respect to international engagement, shaped by three factors. Firstly, the limited impact of the global financial crisis on the Chinese economy increased Chinese confidence in their ability to deal
with the West. Secondly, frustration at what was considered to be attempts by the US to contain China’s rise to great power status. Finally, the uncertainty of Chinese leaders regarding continued economic growth and the potential domestic consequences. This increased assertiveness has been most evident in the approach taken by China in its territorial disputes with regional neighbours in the East China and South China Seas. The provocative declaration of the air defence identification zone (ADIZ) in the East China Sea around the Senkaku Islands in November 2013 and the Nine-Dash Line claims in the South China Seas, have generated renewed uncertainty in the region regarding China’s future intentions, and casts doubt on the sincerity of China’s peaceful development. In light of this recent assertiveness, questions again emerge regarding China’s strategic motivation in the region and the objectives of its foreign policy. Is this renewed territorial assertiveness challenging US regional hegemony, or is it external policy motivated to satisfy domestic objectives?

An important aspect of China’s economic success over the past three decades has been the corresponding development of Chinese foreign policy and the approach China has taken with engaging the international system. By contrast to the isolation-era of Mao Zedong and the formative years under Deng, China’s evolving approach to bilateral relations, multilateral organizations, and security issues has reflected an increasing ‘flexibility and sophistication’ in the execution of foreign policy. What is evident in looking at this evolution however is that China’s engagement regionally and internationally has been largely motivated by domestic considerations, in particular economic and social development. Furthermore, as domestic circumstances have changed, China’s approach to foreign policy has often re-oriented. Although China’s attitude and actions are increasingly more assertive toward regional neighbours, Chinese foreign policy appears to remain focused on the realisation of core interest objectives, as opposed to greater aspirations of global influence. Although territorial claims in the South China and East China Seas are potential drivers of regional instability, which in itself has the potential to retard social and economic development, assertive foreign policy behaviour appeals to a strong Chinese nationalism. Sating the appetite of Chinese nationalism feeds the legitimacy of China’s political system and thus supports the preservation of the CCP.

The use of international engagement to achieve its core interests has contributed to somewhat of a foreign policy dilemma for the CCP. The gulf between the outside world’s perceptions of China as a rising power and the preoccupation of Chinese leaders with internal problems complicates attempts to understand China’s foreign policy, with significant risk of escalating regional geopolitical tension. A relatively minor confrontation in the East China or South China Seas could quickly escalate into a ‘hot’ conflict, particularly against the backdrop of China’s rapid and opaque military growth. However the use of foreign policy to achieve internalised core interests, demonstrates a willingness on China’s behalf to participate in the regional status quo.

**Geo-political tension**

Although the emergence of China has had tremendous economic impact on the region and the wider international community, China’s rise as a great power generates uncertainty regarding her future intentions. History demonstrates that when great powers emerge, they use their power militarily to influence. China’s history is also a useful perspective on the use of force to achieve diplomatic outcomes. Of particular concern to peace and stability in Asia is the increased geopolitical competition between China and the US, and other regional countries. Will a future China be peaceful or disruptive to the Asia-Pacific region?

**Military growth – modernisation or militarisation?**

One clear dividend of China’s economic development has undoubtedly been the capacity to develop and modernise the People’s Liberation Army (PLA). Backed by this economic success, China
has embarked on a rapid modernisation program with a clear emphasis on enhancing naval and missile capabilities and capacity.\(^2\) China accounts for 32.5 percent of military spending in Northeast, Southeast and South Asia (excluding Russia, Mongolia and Central Asia), compared to 18.9, 9.2 and 3.1 percent respectively for Japan, South Korea and Singapore. Defence spending has increased at rates above GDP growth, and the 2012 defence budget was estimated at US$136.7 billion, or about 13.3 percent of the central government’s 2012 budget.\(^2\) Since 1989, China has maintained an average defence expenditure growth rate of 12.5 percent.\(^2\)

The fact China has embarked on such an extensive period of sustained military growth, in a region that has been without war for nearly fifty years, and in the absence of any obvious threat to China, is a cause for consternation. What is the purpose of Beijing’s military growth? Is it the act of an increasingly developed country modernising its military capability as ‘a natural and legitimate outcome of its economic growth’, or is it a deliberate program of militarisation designed to underwrite an increasingly hostile strategic outlook towards the Asia-Pacific region?\(^2\)

This sustained military growth is a factor fuelling regional uncertainty over China’s long-term intentions in the Asia-Pacific. In particular, the development and/or acquisition of capabilities designed for long-range power projection is a source of significant alarm for regional neighbours Japan, Korea and India. All three countries have undertaken modernisation programs of their own during the same timeframe, with India achieving nuclear status in 1998. Perhaps a more significant reaction has been Japanese re-appraisal of its pacifist approach to international relations and the constitutional limitations on the capabilities of the Japanese Self-Defense Force (JSDF). As a result of Chinese military growth, Japan has enhanced security relations with the US and other democratic states in the region in an attempt to hedge against a potentially more hostile Chinese strategic outlook.\(^3\)

Conversely, the development of anti-access/area-denial (A2/AD) capabilities such as advanced submarines, mines, ballistic missiles, cyber and other network disruption enhancements designed to restrict or deny access by a third party to China’s maritime periphery, is a direct threat to the US-led security order in the Asia-Pacific. This A2/AD strategy is designed to deter US intervention in any potential regional conflict in the Taiwan Straits, East China or South China Sea. It is also directly or indirectly, a threat to the interests of all maritime countries in the region.\(^3\)

The lack of transparency surrounding Chinese military development, and China’s antagonistic approach to territorial disputes adds further to the uncertainty over Chinese ambition and objectives within the region. The dispute with Japan over the Senkaku Islands and the universal declaration by China of an ADIZ in November 2013 has increased tensions with Japan. China’s territorial claim of the entire South China Sea and identification of it as one of the country’s core interests constitutes a considerable threat to the national interests of Vietnam, Philippines, Malaysia, Brunei and Indonesia, all who exert sovereignty claims over islands contested by China. As a consequence, these countries are increasingly seeking closer military and strategic relations with the US, raising the strategic stakes in the region.\(^3\)

In relative terms, China is no match for US military capability, or for that matter, the technologically modern JSDF. The refurbished former Soviet Union aircraft carrier Liaoning, a conventionally powered vessel, is the first of four aircraft carriers China will build by 2020 as part of a new “blue-water” navy. Although the subsequent vessels will no doubt be significantly superior to the Liaoning, they would still be no match for the 11 nuclear powered carriers of the US Navy, each capable of carrying twice as many aircraft. Although the PLA Navy is likely to overtake the US Navy in terms of size by 2020, the US has half a century lead on technology and operational experience. China’s “blue water” ambitions are unlikely to pose any threat to America’s Naval dominance. The increased sophistication of China’s asymmetric and A2/AD capabilities, such as advanced anti-ship missiles and
By contrast, Fravel argues that despite growing new offensive air and naval power projection capabilities, ‘China pursues a defensive approach’ to maintaining territorial integrity, and that the primary force structure determinant remains the frontier defence of its land borders. Almost half of the PLA’s combat forces and a sizeable share of land-based airpower are assigned to this objective. Although Taiwan remains the primary strategic contingency task, the capacity of the PLA to resource this mission is, and will continue to be, constrained by the requirements of frontier defence. Furthermore, modernisation is an increasingly expensive undertaking. China’s main objective has been to modernise obsolete forces to a point where they are sufficiently capable enough to deter US intervention in the region on behalf of an ally. The more China commits to sophisticated weapon systems, the more expensive is the cost of sustaining, maintaining and interfacing these capabilities. Training and retention of professional and technically adept forces, in competition with national workforce demands, means that up to 45 percent of the budget is absorbed in personnel costs. In the face of increasing domestic and social challenges, defence growth will be considerably more difficult to justify, and is likely to remain a lower priority overall than economic development. However, continued investment in asymmetric and A2/AD capabilities is a clear indication of Chinese intent to exercise greater influence over the region and possibly contest the US security order, in particular where strategic objectives of neighbouring countries compete with China’s core interests.

United States-Sino relations

President Nixon’s visit to China in 1972 sounded the end of the Cold War in Asia. The US abandoned its position that the Republic of China government on Taiwan represented China, and set the conditions for the establishment of diplomatic relations and formal recognition of the People’s Republic of China in 1979. Following this recognition, the US enjoyed a ‘honeymoon period’ with China, as China sought to draw closer to America to hedge the very real potential for confrontation with the Soviet Union. The Tiananmen Square massacre in 1989, however, demonstrated the enormous gulf between a democratic, values-based US and an authoritarian dictatorship. With the collapse of the Soviet Union in 1989, the rationale for the relationship disappeared. Buoyed by this increased strategic security, China turned its attention to maritime East Asia, pursuing territorial claims and adopting a more aggressive and confrontational approach to Japan and Taiwan.

China is perceived as a potential threat to US national security for three reasons: the country’s geo-strategic significance and growing national capabilities; its expanded involvement in and influence over the international community; and unique historical and cultural factors that with continued growth in Chinese power and influence, could manifest as a direct challenge to the position of the US in the Asia-Pacific. Despite Chinese assurances of their peaceful intentions, the US remains deeply suspicious of Chinese ambition in the region and on the international stage. This has given rise to the argument for the US to seek to contain Chinese growth to ensure no threat to the American balance of power in the region.

In addition to security threats, intensive economic competition between the two countries is a growing concern in some US circles as a threat to American hegemony, and a barrier to bilateral relations. In 2011, China held almost $1 trillion US government bonds, but lagged far behind other Asian and European countries when it came to direct investment. The increasing US trade deficit with China is also a concern with respect to loss of US jobs to China. A major factor in the imbalance has been the undervaluation of the Renminbi and the artificial lowering of the cost of Chinese exports, a factor denied by the Chinese. Although China has recently instituted a range of financial
reforms and the Renminbi is now traded in global currency markets, the threat of increased Chinese economic power and its consequences on American global influence remains prevalent.

Despite these perceived threats, the Sino-US relationship has transpired into one of mutual interdependence. Economically, the two countries have developed very close ties, and this relationship is a ‘broad and strong foundation for bilateral relations’. From a security and stability perspective, both countries also share areas of common concern and expectation, particularly in the areas of anti-terrorism and proliferation of weapons of mass destruction. Both nations have strong incentives to avoid military confrontation.

Regional tension

The most prominent factor in Asia-Pacific tension is the reality that Chinese strategic intentions in the region directly or indirectly affect the strategic interests of the US and every maritime country in East Asia and Southeast Asia. The US is one of the most important military powers in the Asia-Pacific and has bilateral security relations with Japan, South Korea, the Philippines, Taiwan and Australia among others. America has recently also been growing its political and military relationships with Vietnam. The tension between the traditional US alliance system and partner relationships in the region, and the Chinese pursuit of its core interests, in particular Chinese maritime territorial claims, is a constant source of potential military conflict.

Since World War Two, the Asian security framework has been characterised by US pre-eminence. As a result of US security relationships and forward presence, the Asia-Pacific region has remained stable and conflict-free. Ironically, it has been within this stable and peaceful security environment that China has achieved its rapid economic development and military modernisation. There are two reasons why America has been the preferred security hegemon in the Asia-Pacific over China. Firstly, the US has no territorial claims or disputes with countries in the region. Secondly, the US is acutely aware of regional perceptions of its maritime power and seeks regional acquiescence of its naval operations and security relations in general. Consequently, US military presence and its regional bilateral security relationships ‘enjoy widespread support and legitimacy as a stabilising arrangement in the region’. The allies and partners of Washington in the Asia-Pacific remain committed to the regional security role played by the US, and retain significant emphasis on their respective bilateral security relations.

China’s increasing economic and military power however is changing the distribution of power in Asia and there is the likelihood that, eventually, China will become more assertive in challenging US strategic primacy in the region. Hugh White argues that ‘a new order will emerge’ and that a more powerful China will seek to exercise more influence in the Asia-Pacific. The genuine concern within the region is whether China tries to exercise power in ways that are incompatible with the interests and wellbeing of other countries in the region. Tang argues that despite the US remaining the dominant military power in the region, there is a gulf between US Asia-Pacific security goals and its ability to secure those goals. This is particularly so with respect to Taiwan and the South China Sea dispute, where both China and the US view the security of the Sea Lines of Communication (SLOC) to be within their national interest. In this light, there is a compelling argument for the US and China to seek a collaborative, if not collegiate approach, to jointly manage the political and security situation of the region.

Rebalancing to Asia

There is a popular belief that China’s emergence into and eventual shaping of the international system, will occur simultaneously with the decline of US international dominance and a retraction of US interest in the Asia-Pacific. Although the uncertainty surrounding the future intentions of China
towards the Asia-Pacific is a major concern for many countries, an equally major concern in the region is the potential for American disengagement from the region.

When President Obama addressed the Australian Parliament in November 2011 and stated ‘I have made a deliberate and strategic decision, as a Pacific nation the US will play a larger and stronger role in shaping this region’, he announced the US was undertaking a major foreign policy shift to give greater emphasis on security and economic considerations in the Asia-Pacific. Although this policy shift has coincided with some changes to US force disposition in the region, the policy is a multi-dimensional and comprehensive approach to reorient American grand strategy. Aside from changes in US military force structure, the rebalance includes new economic initiatives such as the proposed Trans-Pacific Partnership (TPP) free-trade agreement, the strengthening of diplomatic engagement through closer bilateral relations (including with China), and the deepening of US engagement with regional multilateral institutions.

The “Asia pivot” has been positively received by US allies in the region, who see the rebalancing of US effort away from the Middle East and back to the Asia-Pacific as continued American commitment to the status quo security framework, and strategic reassurance in the face of a rising and increasingly assertive China. To the Chinese however, the announcement, in particular with respect to the military aspects of the policy, has been a cause for alarm. China perceives the pivot as a US attempt ‘to slow down China’s rise and to limit its rightful role as a major regional power’. In Chinese eyes, the public announcement of a return to the Asia-Pacific is a collaborative policy of cold war-style containment by America and her allies, in particular those with maritime territorial disputes with China.

In broader terms, there has been some public misgiving that US rebalancing could increase the potential for conflict between China and the US. Certainly, China’s announcement of the ADIZ in 2013 could be seen as an escalated response in defence of their strategic interests against an increased US military presence. In the context of China’s rise and consequent shift in the distribution of regional power, ‘American efforts to perpetuate primacy in the face of China’s challenge will create not peace and stability but escalating rivalry and a growing risk of conflict’. More recently however, as the US has been placing less emphasis on military aspects of the policy and de-emphasising the focus on China, Chinese reactions have cooled. Nonetheless, the Chinese Premier Wen Jiabao’s November 2012 pledge of China’s support to the Regional Comprehensive Economic Partnership (RCEP), a free-trade agreement between members of ASEAN and six other countries excluding the United States, is seen as a rival initiative to the TPP.

**China’s future intentions?**

The rapid rise of China over the past three decades has already had a substantial effect on the Asia-Pacific region and the world. Its economy has quadrupled in size since the Deng reforms, and China is now the major trading partner for 124 countries. China’s approach to international engagement has also evolved over the same period. Turning away from isolation, China has established a range of bilateral and multi-lateral relationships that now see the country heavily integrated into the institutions of the status quo international system, and playing a major role in the global economy. China’s military capabilities, in particular its missile force, have also rapidly transformed. Its Defence budget has been growing since 1989 by double-digit rates, and the PLA is now the second largest military force in the world. Although regional territorial tensions with Japan, Vietnam and the Philippines continue, and the issue of reintegrating Taiwan back into China remains a simmering source of regional conflict with the US, China maintains its intentions are peaceful and that its new great power strength will never be used aggressively.
Despite these assertions however, the sustained emphasis on military development, together with ‘irredentist territorial claims in Asia’, generates considerable regional anxiety regarding the longer-term intentions of China, and arguably constitutes the most threatening aspect of China’s rise. What are the potential implications for the region and the international system, should China begin to exhibit the atypical attitudes and behaviours of previous emerging great powers, such as 20th Century Japan and Germany? How will China use its military power and how will neighbouring states respond? History demonstrates that where economic power emerges, political and strategic power play usually follows.

Two prominent schools of thought consider the issue of whether or not the rise of China poses a threat, or an opportunity for the international system. “Offensive realism” perceives the rise of China occurring at the expense of US leadership in the region. Backed by the evidence of history, they argue the rise of China will inevitably result in a collision with the existing great powers, with the likely conclusion being intense security competition between the US and China for regional and global influence. According to offensive realists, China will transition its economic might into military power and the US and its allies will respond to constrain or contain this growth. The result will be an adversarial confrontation for power and influence between a declining US and an increasingly stronger China.

On the other hand, “defensive realism” proposes that the rise of China is relative to the international system and that China will not replace the US in world politics. China’s economic growth and therefore future prosperity is inextricably tied to stability within the region. China is also cognisant of the role the US has played in maintaining an unparalleled period of peace and stability in the Asia-Pacific:

‘Increasingly, China’s leaders recognize that economic growth and domestic stability rely on a deepening engagement within the global economy. Regional instability could choke off its vital economic growth, and there is an acceptance among many Chinese policymakers that the American strategic presence is a crucial component of regional security.’

In contrast to offensive realism, defensive realism posits that states seek security, not the maximisation of power, in order to maintain the balance of power. In this context, China will not challenge the US for regional hegemony and will continue to pursue its core objectives through the status quo regional security framework.

An alternative to these two schools of thought is Quansheng Zhao’s managed great power relations (MGPR) model, and the particular relationship between a rising power and a dominant power in world politics. MGPR is a conceptual model designed to comprehend the complex relationship between the US hegemon, and rising China. Zhao argues cooperation by states in international relations is possible so long as three basic elements exist in the relationship. Firstly, both countries must have national interests that overlap, creating mutual benefit in exchange for cooperation. Secondly, sufficient incentives should exist to ensure continued cooperation and to mitigate divergent interests. And finally, effective institutions and mechanisms must exist or be established to facilitate regular engagement. Implicit in this theory, is that if two powers, in this case US and China, can find accommodating positions that allow the two countries to cooperate, a third “shared power” option is a viable course of action.

Internal challenges to future economic prosperity

China’s stunning economic achievements and rapid modernisation of its military over the past three decades have been the catalysts for China’s re-emergence into the international system, and elevation to great power status. Academics, analysts and government officials alike are now viewing
the next phase of this remarkable growth in terms of how China will share, dominate or even alter the international system. The assumption behind this analysis and academic debate is that China will continue to sustain strong economic growth. As previously discussed, China’s continued peaceful rise is contingent on peaceable relations with its regional neighbours and the US. External factors and how China adapts to the international system will have a marked influence on China’s future prosperity. Likewise, there are a number of pressing domestic challenges that China must address if it is to avoid the economic decline that might disrupt China’s future prosperity and continued transition to great power status.

There are numerous internal challenges faced by China today that have been the consequence of China’s rapid ascent to economic success. Urbanisation, industrialisation, marketization and globalisation have all played their part to transform China from a post-revolutionary agrarian society into a modern country. All however have left legacies that today pose serious challenges to continued economic growth. The domestic challenges faced by Chinese leadership are too numerous to cover, and encompass social, political, economic, financial, minority, ethnic, and demographic dimensions. However, four issues that will have significant impact on China’s social cohesion and future economic prosperity are the effects of changing demographics, the rising middle-class, economic disparity, and the environment.

The most pressing demographic challenge for China is its aging population. The “One-Child Policy” when instituted by Deng in 1978 was designed to manage China’s population growth downwards to a more sustainable level. According to some estimates, this policy has prevented China from having a population of over two billion people by 2015. An unintended consequence however has been the acceleration towards a declining population. Unlike Western economies, whose populations did not start to decline until well after reaching a developed stage, China’s population is aging while it is still developing. Furthermore, the one-child policy has also been attributed to an imbalance in the male-female ratio, as a result of the large number of pre-term abortions of female foetuses. A growing number of men will be unable to find a partner, further accelerating the population decline and potentially representing a threat to social stability. Although the liberalisation of the one-child policy announced in November 2013 will likely see a modest increase in the birth rate, it will not address the decline and a smaller workforce will pose significant challenges for sustaining strong economic growth. Compounding the aging problem are two further factors: the lack of an effective pension scheme for retired workers, and insufficient hospital and aged care facilities.

One of the most significant achievements of China’s rise has been the lifting of millions of its people out of poverty and into the middle class. In 2006 it was estimated that the middle class was increasing by one percent per year since 1999, and that by 2020 it will constitute 35 percent of the population. In 2001, contrary to traditional Leninist ideology, the CCP lifted previous exclusions to its membership and permitted private entrepreneurs to join the Party. This decision reflected the very acute understanding by China’s leadership of the relationship between economic performance and private enterprise, and the fact that productivity and prosperity had become the criteria for economic success. Historically, the emergence of a middle class in authoritarian societies has often been a driver of political reform and the gradual establishment of democratic institutions. This is largely due to the fact that as the middle class grows and becomes more prosperous, it increasingly wants to be more participative in the actions of government. The evidence suggests that presently, the Chinese middle class is content with the current political system. Whether or not this continues to remain the case, with or without the endorsement of the CCP, the middle class will influence the political and social future of China.

Despite the enormous rise in the prosperity of China’s middle class, China does not score well comparatively with other countries in terms of per capita GDP growth. In part consequence of the rapid rise of the middle class, but also as a result of the economic approach taken to liberalising the
market, there has been an increasing disparity between regional prosperity, urban and rural opportunity, and the rich and poor.\textsuperscript{69} Once one of the most egalitarian countries in the world, China is now one of the more unequal societies in the developing world.\textsuperscript{70} The rising gulf between the privileged and the underprivileged is a further source of discontent and domestic instability.

Environmental and ecological challenges are also serious challenges for the CCP. China's economic growth has been fuelled by an enormous demand for carbon-based energy sources. China's exploitable energy and water resources are well below the world average in per capita terms, and China has been dependent upon imports to meet this demand. However serious pollution, resource inefficiency and limited use of recycling, challenges the ability of China to sustain strong economic growth.\textsuperscript{71} Combined with 'numerous environmental disasters including flooding, desertification water scarcity and dwindling forest resources\textsuperscript{72}', China's environmental and ecological challenges are another considerable impediment to sustainable economic growth, as well as a major concern for public health and wellness. Unresolved, the environment could also become a driver of social and political dissent against the CCP.\textsuperscript{72}

The magnitude of China's domestic and internal challenges is clearly articulated in the iteration of the "multiplication and division problems". Social and economic problems become a large-scale problem when multiplied by the population. Conversely, any economic, finance or development gain is softened when divided by the same population number.\textsuperscript{73} In recognition of the gravity of these challenges, and to offset their impact on future economic development, the Chinese leadership decreed domestic issues, as opposed to foreign affairs, would be the priority for the Chinese government.\textsuperscript{74} As an outcome of the Third Plenary of the 18th Communist Party Central Committee in November 2013, Xi Jinping announced a range of domestic reforms designed to underpin continued reform and development. These measures included liberalisation of the Hukou system, a relaxation of the one-child policy, liberalisation of the financial system and the full marketisation of the Chinese economy.\textsuperscript{75}

The common thread of all the internal challenges facing China today is the possibility that these problems may, in time, foment the seeds of instability and political dissent. Demographic challenges such as the imbalance between males and females, social problems such as rising unemployment and insufficient health and aged care provision, and the effects of environmental degradation on health, life-style and well being, are all potential sources of social dissatisfaction and are therefore potential threats to the preservation of the CCP. The key performance indicator for the success and justification of the CCP over other forms of government has been economic growth. Ironically, it has been the pace of economic growth that has generated the domestic challenges the CCP now faces.

If the future of the Chinese political system is to be assured, the CCP leadership must address these potential drivers of internal stability. Fundamental to resolving these problems and, therefore by association, preserving the future of the CCP, is continued economic growth. As discussed however, sustaining economic growth is dependant on maintaining regional peace and stability to safeguard Chinese trade. In this light, China's military policies, and their associated regional consequences, appear contradictory to China's economic objectives and, if taken to its logical conclusion, will undermine economic prosperity.\textsuperscript{76} To this end, continued regional and international stability is a vital pre-condition for realising China's domestic national interests.

However, there is a potentially dangerous aspect to China's pre-occupation with internal issues they may inadvertently increase the risk of regional instability. As stated previously, there are a number of sensitive and potentially volatile territorial disputes in the region that will need a focussed and comprehensive foreign policy approach if they are to be resolved or, at best, diffused. The pre-occupation of China's leadership on domestic concerns could obscure the focus on foreign policy development, increasing the risk that China's response to complex diplomatic and foreign policy
challenges may be reactive. This is particularly worrisome in the event of a maritime or aerial incident in the disputed island territories, where a situation could escalate very quickly, unless appropriate response measures are built into China’s engagement strategy. Inattention or ambivalence of China’s leadership to developments in foreign policy reduces the margin for compromise after an incident occurs and increases the risk of regional instability. On the other hand, an unreasonable or escalatory response could be executed to appeal to domestic nationalist sentiment as a means to divert attention away from perceived failings of the CCP. Either way, inattention to foreign policy is a considerable risk to regional stability, without which China will lack the means to achieve economic growth.

What does China want?

In an attempt to get other countries to want what China wants, Zheng Bijian launched the ‘China Dream’ in April 2006, promoting three powerful ideas: economic development, political sovereignty and international law. In November 2012, Xi Jinping, Secretary General of the CCP and soon President of the People’s Republic of China, articulated his vision of “Chinese Dream”. This vision describes both national and personal aspirations for a future China, with its purpose being the reclamation of Chinese national pride and the achievement of personal well being for the Chinese people. In reality, the “Chinese Dream” is China’s strategy to counter the domination of the West’s and, in particular America’s leverage of soft power, to shape the international system. Just as Deng adopted “opening up and reform” and Hu Jintao the “peaceful development” narrative, Xi Jinping has adopted the “Chinese Dream” as his hallmark for the next phase of China’s rise.

Jakobson argues that China, by in large, seeks respect and to be treated as a great power with influence within the international system. Xi Jinping’s “Chinese Dream” illuminates the aspiration of a return to great power status. The vision is driven by acute nationalist desires to right the wrongs of Western imperialism in the 19th century, and redress the humiliation of China’s domination by the West during the 19th and first half of the 20th centuries. In this historical context, there is some conjecture that China may be abandoning its peaceful development narrative in favour of a more adversarial “realist” approach. However, China has made clear the desire for this return to be a peaceful process, being only too well aware of the consequences of Germany and Japan’s great power ambitions in the 20th century.

China’s marketisation strategy and the internationalisation of the Chinese economy have established interdependence with China and the rest of the world. China is fully integrated into the international system and benefits from continued stability of the status quo. The economic and social benefits of this integration would be significantly diminished in the event of conflict. ‘The economic motor must be kept running at all cost, an impossible feat without stable relations with its Asian neighbours and the global community’. China is on the threshold of superpower status however, as yet, still lacks the experience to apply its influence. ‘Herculean in ambition, but still brittle’ politically, socially and economically, China’s struggle is to confront the challenges of sustainable economic growth, while continuing to progress sovereignty issues and national security. Despite their insecurities, Chinese are pragmatic and are well aware that their interests and future national ambitions are intimately inter-twined with the US and the western system. Furthermore, stability, both regionally and internally is fundamental. The concept of “gradualism” is, therefore, deeply rooted in the way China acts, both in domestic economic matters and international relations. As China finds its way as a new power, it will test the water but not rock the boat. To this end, although the US-led security framework in the region is less than ideal, it is the paradigm within which China must cooperate in order to maintain the peace and stability necessary for continued prosperity.

China is also acutely aware of history, both in the sense of its view of national grievances, but also in the perspective of the international system. China appears genuinely motivated to not follow the
destructive path of 20th century great power transitions, such as Germany and Japan. Although Chinese foreign policy will no doubt continue to sometimes clumsily threaten neighbours, China remains deeply aware, and responsive to, the perception of other countries in the region. Furthermore, Chinese foreign policy will continue to be largely driven by domestic considerations, with the most pressing of these being resolution of the aforementioned internal problems. Unchecked, they remain serious impediments to economic growth and drivers of social and political instability. Unless threatened, China is unlikely to challenge the status quo of the international system. On the contrary, as a rising power within the Asia-Pacific, China has portrayed itself as not only an engine of economic growth but as a stabilising force for the region.\textsuperscript{83}

Furthermore, the success of China over the past three decades has occurred within the context of a peaceful environment. The Chinese economy has been the beneficiary of the ‘post Cold War reduction in inter-state violence’ and the US-led regional security order in the Asia-Pacific. Continued economic success, which by and large has been the single measure of the CCP’s performance, is contingent on regional and international stability.\textsuperscript{84} Economic growth is also essential for addressing the domestic challenges that may become future drivers of social instability. Continued economic and social development is at the heart of the preservation of the CCP. Accordingly, regional and international stability, and peaceable relations with the US and regional neighbours, is a paramount objective of the Chinese government.

Shared power relations?

Despite the many differences between the US and China, both countries continue to share many mutual interests that precipitate peaceable relations and a stable international environment. Both are significantly integrated into the others economies. In 2013, China was second only to Canada in overall total US trade. Although the trade balance deficit of -24.5 percent is in China’s favour, the importance to the US of the Chinese market cannot be understated.\textsuperscript{85} Both countries also have vested interests in the global commons, including the SLOC and space. They also share similar global concerns, such as terrorism, nuclear proliferation, energy security and public health, where a cooperative and collaborative approach serves mutually beneficial purposes. Cooperation between the two countries has also demonstrated some success in diffusing regional tensions such as in the Taiwan Straits, potentially the most dangerous flashpoint in the Asia-Pacific region. The US maintains a strategy of ‘peaceful unification’, opposing any unilateral decision by either China or Taiwan of altering the status quo. The willingness demonstrated by the US in 2004 to co-manage this dangerous regional flashpoint and prevent confrontation with China over the issue, suggests the US is accommodating of a collaborative approach to maintaining stability in the region.\textsuperscript{86} Despite challenges and perceived threats in China’s foreign policy, there remain significantly more opportunities and incentives for both countries to maintain a stable and constructive partnership in the Asia-Pacific region.

Implications for Australia: Balancing United States and Chinese relations

‘The rise of China has constituted the most significant and consequential geopolitical shift since the end of the cold war’.\textsuperscript{87} Economic and military competition between the US and China over influence within the Asia-Pacific has significant implications and potentially dire consequences for Australia. On the other hand, the continued economic rise of China presents Australia with a range of economic opportunities. China is Australia’s number one trade partner, accounting for 33 percent of GDP in 2014. The continued peaceful rise of China will, accordingly, be to Australia’s benefit. However, the rapid modernisation of the Chinese military, together with China’s belligerent actions over its territorial disputes, has the potential for either direct or indirect conflict between the US and China. Conflict in East Asia would not only threaten Australian economic prosperity, but might also create the situation where Australia’s economic relationship with China could be incongruent with
its security alliance with America. How the Sino-US relationship evolves for better or otherwise will be a significant factor in balancing Australia's relationship with both countries. The imperative for maintaining regional harmony and mitigating the potential for this relationship dilemma will be the manner in which China either conforms with, or is permitted to participate in, the regional security order.

**Three future scenarios?**

In line with the offensive realist thought, China will increasingly seek to challenge US hegemony in the Asia-Pacific in order to fulfil its strategic objectives. Continued Chinese economic and military growth, together with its growing maritime strategic orientation and irredentist sovereignty claims in the South China Sea and the Japanese Senkaku Islands, leads to the possibility of a redefined strategic outlook both regionally and globally. ‘A new period when the West ceases to manipulate China and discovers that China will dictate her own rules to friends and partners’. This redefined outlook could have potentially threatening implications for US strategic interests, both in terms of hegemonic influence in the Asia-Pacific, but also more broadly as the unipolar superpower in global affairs.

Alternatively, the continued rise of China, both militarily and economically, could continue within the status quo of the current US-led regional security order and the contemporary international system. Furthermore, China’s continued emergence into the status quo as a great power could also occur concurrently with the controlled transformation of China’s political system by the CCP, to a ‘democracy with Chinese characteristics’, completing the evolution of reform started in the 1980’s. This evolution could occur within the status quo context of US leadership of the international system or, subject to the US approach to China’s rise as a great power, within a bipolar or, indeed, multipolar context.

A third, and potentially more challenging scenario, is the failure of China to address challenging domestic social and political issues, resulting in the slowdown or reversal of economic growth. Increased social and political instability inside China could lead to the emergence of a ‘weak, insecure and defensive Chinese regime’ with equally challenging implications for the US and its Allies. In this scenario, a collapse of the current Chinese political system could result in internal instability and division, with devastating implications for the Chinese and regional economies. Furthermore, as a means of survival the CCP ‘could become more belligerent and assertive over critical nationalist issues’ such as Taiwan, Japan and the South China Sea, raising the prospects of serious conflict in the region with similarly devastating economic consequences.

**Engaging China in the status quo**

It appears highly unlikely, when looking at the three possible future scenarios of Chinese regional behaviour, that China will, in the foreseeable future, mimic the aggressive and power challenging behaviours posited by the offensive realists. As we have seen in the previous sections, future Chinese power will be fuelled by continued and sustainable strong economic growth and this growth depends on ongoing access to international resources and markets. The disruption of either would jeopardise economic growth, leading to internal instability that would threaten the CCP. Chinese economic growth and internal stability, therefore, demands a stable and peaceful region. China’s substantial integration into the international system, in particular her increasing willingness to engage through multilateral institutions, and the interdependence of China’s economy with the global economy, suggests an acceptance if not dependence on the status quo. Although regional tension will continue to exist as a result of Chinese foreign policy, this will largely be the result of China testing the water, rather than deliberate confrontation. However, relatively minor incidents could be the genesis for major confrontation, unless there is open and constructive engagement.
The potential for ‘conflict of interest’ between Australia’s economic interdependence with China and her security alliance with the US is a consideration for the Sino-Australian bilateral relationship. Australia and China have had formal relations for over 78 years and re-established diplomatic relations in 1972. Australia regards its relationship with China as one of great strategic importance. Australia has long sought to engage China within the status quo, and an important objective of Australia’s engagement with China has been to ‘bind’ the country ‘into a complex web’ of multilateral frameworks to integrate it economically to the region. Australia has been successful in this regard and as a result China engages regionally through three cooperation frameworks: the Asia-Pacific Economic Cooperation (APEC) established by Australia in 1989; the Association of Southeast Asian Nations Regional Forum (ARF); and the East Asian Summit (EAS). The APEC forum facilitates regional trade liberalisation and economic technology cooperation and includes Chinese Hong Kong and Chinese Taipai.

China, however, continues to remain indisposed to engaging in multilateral forum where the US is eminent and, in fact, seeks to establish its own multilateral alternatives to arrangements that form part of the US-led security framework. Australia must continue to positively contribute to integrating China and the US into the regional security and economic forum, in order to foster continued Chinese participation in the status quo. Although the US and China relationship is complex and reflects two different and sometimes opposing perspectives, continued bilateral and multilateral engagement and dialogue is necessary to generate mutually beneficial economic competition, and maintain a peaceful and stable Asia-Pacific.

Nonetheless, a critical part of coaxing China’s continued embrace of the status quo is being able to convey differences of opinion both ways. Australia needs to continue to be sympathetic to China’s concerns, but at the same time must be prepared to make clear her position when Chinese action is counter to the norms of the status quo. China might not like that Australia sometimes vocalises its disagreements publicly but accepts that, as a democratic country, this remains a core value. Australia’s public criticism of China’s announcement of the ADIZ in the East China Sea was justifiable and necessary to reflect this position.

Moderating US attitudes to China

The Australian dilemma is that its traditional close security relationship with the US may in the future complicate, and perhaps irritate, its increasingly close ties with China. Conversely, Australia’s economic relationship with China, no matter how interdependent, runs the risk of being compromised in the event of US-Sino conflict. The management of the relationship between Washington and Beijing will be of paramount importance for strategic stability in the Asia-Pacific region. Australia must balance its responsibilities as an ally of the US and, at the same time, avoid provoking China and disrupting the Sino-Australian economic relationship.

As mentioned above there remain many issues on which China and US do not agree, or share different perspectives: human rights and religious freedoms, trade practices and competition, political incompatibility, economic competition and militarisation. Addressing these differences and the issues in East Asia where Chinese and US interests are divergent requires cooperation with China, as opposed to competition. Tang proposes a new collegiate approach to regional leadership, between US, China and Japan. Given Australia’s strong bilateral relations with both countries, Australia is well placed to leverage the trust and mutually beneficial aspects of each relationship to support this concept. Facilitating Chinese continued acceptance of the status quo requires some give on the behalf of the US, as well as Chinese conformance with the norm. Accordingly, Australia must continue to exercise frank engagement with the US to highlight the differing views in the relationship, and encourage the US to engage with China in a manner that accommodates a greater
involvement of China in regional security. Although Chinese foreign policy inconsistencies must continue to be challenged where they upset the regional security balance, the US must remain aware of the pressing internal challenges to China’s future growth and ensure reaction to possible Chinese foreign policy ‘short-sightedness’ is not escalatory. US and regional reaction should be firm in preserving international convention but not provocative, particularly where responses may be seen by the Chinese as humiliating or condescending.

Effective relations between China and the US, and indeed the other regional powerhouses Japan and South Korea, is essential to regional security and, therefore, to Australian prosperity. Balanced, direct, open and impartial relationships between Australia and these countries are paramount to facilitating closer alignment on issues of disagreement and in maintaining regional harmony. Moderating the US position on many of these issues is fundamental to achieving the compromise that will enable a more collegiate approach to security and stability in the Asia-Pacific, perhaps along the lines of the MGPR model proposed by Tang.

Conclusion

The rise of China and its impact on the international system has been a subject of prolific debate among academic and security strategists over the past decade. Much of what has been written regarding Chinese motivation has been sensational and predictable. The reality is that China already exercises significant leverage over the East Asian region through trade, investment, development and, more recently, through an increasingly participative approach to the international system. Economic growth is essential for the next stage of China’s rise, in particular to address pressing domestic issues that will be barriers to future economic prosperity. A stable, peaceful and integrated regional environment is critical for China achieving its economic objectives. Whether or not China’s long-term objective is to challenge the US for global hegemony is a matter for speculation, and certainly unlikely to be an eventuality in this century. However, for the foreseeable future, China needs a stable Asia-Pacific, good relations with regional neighbours, and cooperative and working relations with the US.

The economic ascent of China has provided significant economic benefit and prosperity to the region and, in particular, Australia. Since Deng’s “reform and opening up”, China has become increasingly more participative in the international system through trade and investment. This has created interdependence between China, the international economy and many regional and non-regional states, including Australia. Economic capacity has, in turn, enabled military modernisation and increasing relative power has instilled more confidence in pursuing a greater role in regional affairs. It has also created unfavourable consequences however that must be addressed to ensure continued prosperity.

In the exercise of foreign policy, China’s primary motivation remains driven by domestic considerations, namely economic development, but is also aimed at nationalist sentiment. The preservation of the CCP is a core interest for Chinese leaders and foreign policy, poorly executed or not, is a vehicle for achieving this objective. In this respect, China is leveraging its economic dominance and strength within the region to pursue historical territorial claims as part of the new Chinese Dream. This vision makes clear China’s intention of redressing the humiliation at the hands of the West in the 19th century, and to once again become a wealthy and strong power. The dream similarly carries a vision not dissimilar to the “American Dream”, in that it also seeks to provide for the wellness of its citizens, achieving a Chinese way of life that is conducive to peace and stability. Challenging this dream however is the ability of the CCP to address the extensive domestic challenges to future economic growth, which if not resolved, have the potential to destabilise the country and the region. Furthermore, incongruent military policies and a preoccupation with domestic issues at the expense of a longer-term view of foreign policy, continues to create tension.
and increase the risk of regional conflict detrimental to Chinese and regional prosperity. China’s paradox however is that failure to address these issues could very well create the conditions for instability and economic regression that, in turn, could result in similar regional consequences.

The most dangerous outcome for Australia in terms of the future of China is the offensive realist scenario, where China seeks to use its military power to become the security and economic hegemon of the Asia-Pacific. Given China’s economic interdependence with Japan, South Korea and the United States, however, and the collaborative effects of globalisation, this option is potentially the least likely in the next 20 years. China knows that to address the domestic challenges necessary to guarantee long-term and sustainable prosperity, it requires peace and stability within the region and, in particular, with the US. However, any perceived threat by China to constrain or contain their rise, or to deny their rightful participation in the Asia-Pacific security order, could precipitate a revisionist approach by China, and set the conditions for a dangerous escalation of strategic rivalry and competition. Consequently, although the preferred outcome for Australia is the continuation of US strategic influence in the region and the continued participative approach by China in the status quo, an alternative solution of agreed US and Chinese power sharing also serves Australian and regional interests. If this is to become a viable course of action, however, Australia and other regional nations must continue to work with the US to ensure acceptance of China as an equal partner.

Endnotes

6 Cohen, p. 699.
9 China joined APEC in 1989 as a founding member and agreed to the admission of Taiwan and Hong Kong in 1991 (see Dittmer, p. 919).
12 Waldron, p. 715.
16 Kitano, p. 39.
17 ibid, pp. 41-44
18 Medeiros and Fravel, p. 24
19 Kitano, pp. 46-47.
21 ibid, pp. 107-117.
22 Medeiros and Fravel, p. 24.
23 Linda Jakobson. ‘China’s Foreign Policy Dilemma’. The Lowy Institute For International Policy, (February 2013), p. 3.
26 Suisheng Zhao, p. 105.


30 Waldron, pp. 721-727.


33 Article. ‘China’s Aircraft carrier is Nothing to Fear’. *Bloomberg View*. http://bv.ms/1g9llPg Editorial. 9 April 2014. Last accessed 25/04/14.

34 Fravel, p. 732.


41 Jinghao Zhao, pp. 631-632.

42 Zhao and Liu, pp. 592-593.


44 John Lee, 407-408.


48 Ibid, pp. 81-84.

49 Xiaosong Tang, p. 594.

50 ibid, p. 598.

51 Lenore Taylor. ‘Changing Fortunes Dictate Another Presidential Pivot’. The Sydney Morning Herald,

52 In 2012 a first rotation of US marines occurred in Darwin, with the number expected to increased to a Marine Air Ground Task Force (MAGTF) of 2500 personnel by 2020. In April 2013, the first of 4 new Littoral Support Craft arrived in Singapore as part of the new redistribution of the USN, which will eventually see 60 of the USN based in the Asia-Pacific. A similar force ratio is anticipated for the disposition of the USAF overseas combat capabilities.


54 Jakobsson, p. 5.


57 Hale, p. 27.

58 Waldron, p. 716.

59 Fravel, p. 705.


61 Quansheng Zhao, p. 609.

62 Zhao and Liu, pp. 586-7.

63 Quansheng Zhao, p. 610.


65 Quansheng Zhao, pp. 609 and 623.


67 Hale, pp. 5-6.

79 Whereas previously market forces played a ‘basic’ role in the Chinese economy, they will now be the decisive factor in
Chinese economic activity.
80 Following the 18th Party Congress in 2012.
81 Following the 2012 Party Congress, market forces are now considered the decisive factor in Chinese economic activity.
82 Glasser and Medeiros, p. 298.
83 Following the 18th Party Congress in 2012, the role of market forces in the Chinese economy increased significantly.
84 Following the 2012 Party Congress, market forces are now the decisive factor in Chinese economic activity.
85 Source: United State Census Bureau, Foreign Trade Data. www.census.gov/foreign-trade/statistics/highlights/top/ Last
accessed 02/05/14.
86 Following the 2012 Party Congress, market forces are now the decisive factor in Chinese economic activity.
87 Following the 2012 Party Congress, market forces are now the decisive factor in Chinese economic activity.
88 Page 191
Shale: A fracking geopolitical game changer?

Captain Ian Bisson, Royal Navy

Introduction

There are two shale revolutions ongoing, in oil and gas. Each has different strategic implications, but together they will fundamentally alter the world geo-politically. The revolution has started with increasing production of shale oil and gas in the USA using the process of hydraulic fracturing ("fracking"). The implications of the USA achieving self-sufficiency in energy, a goal of over 40 years, could be significant if the world’s only super-power retreats to a more isolationist foreign policy and security stance. A more profound revolution, with massive economic and geo-political implications for all current energy exporting countries, could follow if the worldwide shale oil and gas reserves prove to be as exploitable as those in the USA. However, there are a number of obstacles, evident from the USA already, that mean, even if shale oil and gas reserves are proved and could be exploited economically, they may be left in the ground. These include environmental concerns, regarding the process of fracking itself and the contamination of water supplies, access to sufficient water and pipeline infrastructure and the capacity of industry to drill and service the large number of wells required.

Background

Hydraulic fracturing, commonly known as fracking, is not a particularly new process but, in conjunction with several different technological advances, it is being used increasingly to extract both oil and gas from previously unexploitable reserves trapped in shale rock. These techniques of horizontal drilling, extending out several miles, multi-stage hydraulic fracturing, specific drilling bits for shale, and improved seismic imaging and enhanced interpretation, have contributed to the rapid growth of shale oil and gas production in the USA. As an example, crude oil production in the USA increased by 790,000 barrels per day (b/d) between 2011 and 2012, the largest increase in annual output since the beginning of USA commercial crude oil production in 1859. The rate of increase has surprised most analysts, including the USA government’s own Energy Information Administration (EIA), which has repeatedly, over the last few years, had to revise up its annual forecasts of shale oil and gas production.

The process of bringing a shale oil or gas well to completion is generally short, taking only 70 to 100 days for a single well, after which the well can be in production for 20 to 40 years. The process for a single horizontal well typically includes four to eight weeks to prepare the site for drilling, four or five weeks of rig work, including casing and cementing and moving all associated auxiliary equipment off the well site before fracturing operations commence, and two to five days for the entire multi-stage fracturing operation. The fracking process itself requires large volumes of water, which is mixed with sand and chemicals, such as guar gum to prevent coagulation of the sand, and which is then pumped at very high pressure to open cracks in the shale rock. The water is then removed, leaving behind the sand which keeps the cracks open, and from which the oil or gas flows out into the well bore and is collected at the surface. Shale oil is usually transported by pipeline (from the larger well sites) or by truck for further distribution by either pipeline or rail to the refinery. Gas is compressed and piped into the gas network. A single well could have up to 10 horizontal wells emanating from it.

Once completed, the production site is reduced to about the size of a two-car garage. Local impacts, such as noise, dust, and land disturbance, are largely confined to the initial phase of development.
Let us set as our national goal, in the spirit of Apollo, with the determination of the Manhattan Project, that by the end of this decade we will have developed the potential to meet our own energy needs without dependency on any foreign sources

(President Nixon) 7

The USA has been a producer of oil for over 100 years and the shale oil revolution is but the latest phase. Conventional oil production peaked in the 1970s and 1980s at 8.8 million b/d9, and then declined until the USA imported more than 65% of its oil needs in 200610. Shale oil extraction has reversed that trend and the EIA expects USA crude oil production to continue rising over the next two years, with output of 7.25 million b/d by end 2013 and to 7.8 million b/d by end of 201411. Most of the USA crude oil production growth will come from shale formations located in North Dakota and Texas; oil production is projected to continue to rise over the rest of the decade, reaching 8.5 million b/d by 2017 and 10.5 million b/d by 203012. Demand is expected to remain constant, or even fall slightly, as improvements in energy efficiency offset the increase in consumption that would otherwise be expected due to a growing economy13.

However, total USA oil consumption remains at around 18.6 million b/d14, therefore the USA continues to be a large net importer of oil, with the majority coming from other North and South American states, particularly Canada (2.6 million b/d), Mexico and Venezuela (about 1 million b/d each). A further 1 million b/d arrives from each of Nigeria and Saudi Arabia with other suppliers making up the balance (of about 3 million b/d)15.

The EIA also projects that Canadian production could grow to 6.6 million b/d by 203516 due to an expansion of output from both shale oil and oil sands, most of which would be exported to the USA through the existing and planned pipeline systems. Mexico and Venezuela are amongst the largest oil producers in the world, but the amount of oil produced by these countries has steadily decreased in recent years due to poor management and lack of investment by the state owned oil companies. Potential reforms could liberalize the sector, promote greater foreign investment and lead to significantly greater exports17. Due to the type of oil18 and its proximity to the USA refineries and market, export to the USA is most likely irrespective of politics.

This increase in supply has not contributed to significantly lower oil prices, because worldwide supply and demand are currently approximately in balance and Saudi Arabia, in particular, pumps more or less oil as required to regulate the price. The increase in USA production in 2012, whilst about 14% of USA supply, was only 1% of worldwide production19, and gains in the USA, Libya and Iraq were offset by reductions in Mexico, Sudan/South Sudan and Iran. In 2013, the situation partly reversed, with strikes and terrorism affecting supplies from both Libya and Iraq; meanwhile Saudi Arabia increased production to compensate. Current predictions are for a gradual, and matching, increase in supply and demand with prices remaining stable around the $100 - $120 per barrel20.

The prospect has been raised of the USA becoming a net oil exporter, or at least not a net importer of crude oil, but this is considered achievable only in the most optimistic forecasts21, and not until 2030 at the earliest. However, even median forecasts of supply and demand show that the North and South America together could be self-sufficient in oil within a few years22, and thus, it has been asserted, the USA could insulate itself from global supply shocks, in particular from the Middle East.

The oil market, however, is more complex, with both crude oil and petroleum products widely transported in order to make best economic use of the available refinery capacity; those refineries
themselves are often optimised to process specific crude oil types. The USA refineries on the Gulf coast have long been specialist in refining heavy, sour crude of the type produced by Saudi Arabia and Venezuela. It is likely that the USA, even if it could source all its oil from the Americas, would continue to import and process oil from countries further afield. Thus, the implications of supply and demand balance self-sufficiency are not so much a cessation of imports from beyond North and South America, but a greater resilience to supply shocks and higher energy security. Even though the USA still imports crude oil, its refineries produce a range of products in excess of internal demand, particularly diesel but also, more recently, petrol; exports of refined petroleum products reached nearly 2 million b/d in September 2013.

The largest direct impacts on the USA from increased shale oil production are the economic impacts: an improvement in the balance of payments (amounting to a net change of $36Bn every year); the renaissance in heavy industry and the railways in the mid west to support the extraction and distribution processes; growth of petro-chemical plants; and an increase in refined petroleum product exports. The last two of these are already showing their effects elsewhere, particularly in Europe, and the UK’s Grangemouth refinery and petro-chemical plant, which nearly shut recently due to USA (and Saudi Arabian) competition, is but one example. The USA has had, and continues to have, a ban on crude oil exports, deriving from its long-standing goal to be self-sufficient, but its refineries and petro-chemical plants have been permitted to sell refined products. A typical petro-chemical plant requires naphtha (from oil) or ethane (from gas) as feedstock to produce the base chemicals for the plastics, solvents and paint industries; the shale oil and gas have provided a competitive advantage for US companies which they have seized. USA refineries have similarly benefitted, and surplus diesel and petrol have found a ready market in Europe where both raw material and environmental costs are much higher. Grangemouth’s petro-chemical plant was only saved by investment to allow low-cost USA shale gas to be imported as feedstock.

The major geo-political impact could be the decreasing significance of the Suez Canal as an oil transit choke-point. In 2012 average daily flows, by both ship and pipeline, were less than 2 million b/d, and Middle East crude oil will increasingly be heading east to Asia. However, even if the USA is insulated from Middle East supply shocks, it would still feel the effects of higher oil prices and the fall-out from the impact on the global trading and financial systems. Thus the USA, irrespective of its own oil supply situation, would remain highly susceptible to any global supply shocks and therefore continue to have a national strategic interest in the level of oil production and its secure distribution worldwide.

World shale oil

There are number of significant reasons why the success of the USA in exploiting its shale oil reserves, and apparently so rapidly, will not be repeated, possibly for many years, in the rest of the world. First, the USA started the shale revolution with a long history of independent, risk taking oil companies, able to exploit a new opportunity and get the financial backing to invest in the equipment. At the start of the shale oil boom, the USA already had a large oil producing industry with over 350,000 active oil wells in 2009 and 400 active rotary oil rigs, but was able to bring around 1,000 rigs back into oil production within 3 years (a majority of which are able to drill horizontally). Second, oil was a foundation industry in some of the mid-West states, therefore, there is both a general acceptance by the population and the shale oil industry is seen as a renaissance of traditional bedrock of USA power and wealth. Third, the USA has the space to accommodate a vast oil exploitation industry, with minimal impact on other land users or population and fourth, it has the large water resources necessary for the extraction process. Fifth, the USA already had an infrastructure of pipelines to both collect conventional oil from the fields in the mid-West, centred on Cushing, Ohio and a distribution centre for refined petroleum
products from the Gulf of Mexico refineries. The large pipeline from the Gulf to Cushing has been reversed to now take crude oil from Cushing to the same refineries. The pipeline network has been supplemented by use of rail, with oil replacing coal deliveries on the same lines. This infrastructure of pipelines, refineries and storage facilities would need to be replicated elsewhere in order to provide the backbone of a shale oil industry. Lastly, in the USA landowners own the sub surface mineral rights whereas in other countries it is the state. This provides a huge incentive for landowners, who often receive payments for access to their land and in compensation for the disruption caused by drilling, with royalty rates of 12.5% to 20% for every barrel extracted. In just one state, Pennsylvania, in 2008, the land acquisition payments alone amounted to $3.5bn.

The EIA estimate that there are 335 to 345 billion barrels of shale oil worldwide, compared to conventional reserves of over 1,400 billion barrels. The largest reserves of shale oil are in Russia, followed by US, China, Argentina and Libya. Whilst total shale oil production will expand by 7.5 million b/d by 2030, accounting for nearly half of the anticipated increase in global oil production, over two-thirds of this growth will be in North America. Only Russia and China, which could match the USA industrially and provide the right financial incentives and regulatory regimes, are expected to develop a significant shale oil sector; but this is still expected to be no more than 1.4 million b/d in Russia and 0.5 million b/d in China. Therefore shale oil is not going to be a significant resource, except for the USA, for at least two decades and the strategic implications of shale oil alone are confined to the national economic boost and the reduced significance of the Suez Canal as an oil transit chokepoint.

USA shale gas

The shale gas revolution in the USA is, however, much greater than that in shale oil, with greater significance for geo-politics. The first production shale gas well, using many of the new techniques, was only drilled in 1998 in Texas. By 2005 there were 15,000 wells producing about 1 trillion cubic feet (tcf) and this increased rapidly so that, by 2011, shale gas production was 9 tcf/y out of total USA natural gas output of 23 tcf/y. Natural gas production in the USA is forecast to increase by 50% by 2040 to 33 tcf/y, with almost all of this increase coming from exploitation of shale formations across the USA. The USA currently imports only about 10% of its gas needs, mostly in the winter months when consumption exceeds production (this is down from an overall 25% as recently as 2005). With gas production rising at about 5% per year, the USA is predicted to become an overall net exporter of gas within 2 years, although there is tension in the USA between the gas producers who seek export opportunities in order to increase their revenue through higher prices, and heavy industry who seek to maintain low gas prices in the USA in order to gain competitive advantage against Asian competitors. By May 2013 only one export licence for gas had been granted but three more have since been granted with a further 20 applications in the pipeline.

The rapid increase, over a relatively short time, in the supply of natural gas from fracking, and which has been significantly greater than the expectations of even the USA’s own EIA, has driven down prices and made USA natural gas much cheaper than anywhere else in the world. In comparison to the world oil market, the gas market is less fungible, because of the way that gas has been mostly transported through fixed pipelines from the suppliers’ gas fields to the consumers’ gas distribution network. Thus the increasing abundance of gas in the USA has resulted in reduced prices in the USA, even whilst gas prices elsewhere in the world are up to four or five times greater. The benchmark price for gas is measured in $ per million British thermal units (mbtu); as recently as 2007, the price for gas produced in the USA, in Germany from Russian supplied gas, from the UK North Sea and in Japan for Liquefied Natural Gas (LNG) was within $1 for each at about $7/mbtu. Following the rapid increase in shale gas in the USA, for which there was only the USA market, prices diverged, with the price driven down to below $3/mbtu; meanwhile
the price of gas in Europe rose to between $8-10/mbtu following the overall increase in energy prices worldwide due to increase demand from Asia, and the price of LNG to Japan peaked at $16-18/mbtu following the Fukushima nuclear reactor disaster.

These relatively low USA natural gas prices, without ready access to export markets to take advantage of higher prices elsewhere, have instead spurred increased use in the industrial and electric power sectors, and natural gas also continues to capture a growing share of total USA electricity generation, from 16% of total generation in 2000 to 24% in 2010. This use of gas in power generation has displaced coal, traditionally the major fuel for power generation in the USA, and this coal has been exported at low prices to Europe where it has undercut the European gas market for power generation. In the UK for example, gas plants have been mothballed as the large coal plants have been used, not least because they also have a limited life before EU emissions targets force them to be shut from 2015.

In the USA, because natural gas combustion can emit less than half as much CO2 as coal combustion, per unit of electricity output, USA carbon emissions have fallen in the last few years. The relatively cheap natural gas has also resulted in a number of manufacturers and heavy power consuming industries investing in new plant in the USA instead of Asia.

Natural gas is expected to continue to take power generation share, mostly from coal but also nuclear, reaching 30% by 2040. Natural gas also reaches other new markets, as a fuel for heavy-duty freight transportation (trucking), and as a feedstock for producing diesel and other fuels, displacing both conventional oil and biofuels. This switch to natural gas or gas-derived fuels is predicted to affect sales of all electric vehicles, with only 100,000 being produced per year by 2030, against 1.3 million hybrid or alternative fuel vehicles.

The USA benchmark price, whilst rising to $5-6 per mbtu by 2020, is predicted to remain relatively low against international energy prices; this will benefit the bulk chemicals industries directly and the metals industries through lower energy input costs. One prediction is for a doubling of the USA chemical industry by 2020, leading to a $250Bn per annum rise in USA manufacturing output. With improved efficiency of energy use and a shift away from coal and oil, USA energy-related carbon dioxide (CO2) emissions should remain more than 5 percent below their 2005 level through to 2040. In this regard, and as seen in some quarters, natural gas is considered to be a bridge to a low carbon future, as an alternative to the current European recourse of very high fossil fuel prices to stimulate renewable energy production. An alternative perspective is that shale gas continues mankind’s reliance on fossil fuels, and it is the renewables sectors that are feeling the greatest impact; the evidence so far, even in the USA, is that gas is firstly supplanting coal for power and oil for transportation and that wind and solar energy investment has not yet been adversely affected.

Liquefied natural gas

It is not possible to consider the impact of shale gas beyond the continental USA without examining the market for gas distribution. Until the 1960s gas was moved between states and within states by an infrastructure of fixed pipelines. In 1964, the UK became the world’s first Liquefied Natural Gas (LNG) importer and Algeria its first exporter. Since then over 200 liquefaction and re-gasification trains have been built worldwide with a capacity to transport about 80 tcf/y. Shale gas and LNG are now complementary aspects of this energy revolution. Shale gas allows exploitation of previously unexploitable reserves and LNG provides the means for more flexible transport than fixed pipelines, thus creating a more fungible, global market for both source of supply and pricing.
The USA shale gas revolution, in shifting the LNG market from regional point-to-point contracts to a truly globally-traded, fungible, commodity already has implications for the traditional gas exporters using their pipeline infrastructure.

The price differential between USA and the rest of the world is driving investment in export facilities. The first cargoes for export from Texas require that a LNG re-gasification train, completed only in 2007 to import LNG from Qatar, is converted to allow gas liquefaction into a LNG carrier. The prospect of USA natural gas entering the market is changing the way natural gas has been traded globally for years; unlike oil where cargoes are bought and sold whilst the ships are at sea, LNG contracts have been long term as they require huge investment in liquefaction plants at the exporting terminal and a regasification plant at the importing terminal to convert LNG into gas for transmission. Traditionally LNG prices were linked to oil prices, but the USA natural gas bonanza is pushing suppliers in other regions to lower prices in some contracts even before the country ships its first cargo of LNG. BP has recently agreed to supply 2 Japanese utilities with natural gas for 15 years at a price partially linked to the USA benchmark natural gas price.

**World shale gas**

Until 2009 projected shale gas reserves were low but by 2011 they were estimated at 50% all other existing reserves (equivalent to 100 years supply at current global consumption rates) at 6,700 tcf; with 28% in North America, 18% in South America, 15% in Africa, 21% in Asia, 6% in Australia and 9% in Europe. On this basis, worldwide shale gas production is predicted to quadruple by 2040, with the USA contributing about 50% of the total production. However, before examining the opportunities for countries, other than the USA, with the most significant exploitable shale gas reserves, whether fracking for gas becomes more widespread depends in these countries as much on the environmental issues as it does on the logistics and economics of extraction.

**Environmental issues**

A discussion on shale oil and gas requires a good understanding of the environmental issues, particularly as arguments about its impact are being used by some countries, especially currently in Europe, to either prevent fracking from taking place or to constrain it by regulation. Fracking has been used 1 million times in the USA over 60 years so there is plenty of evidence that both sides of the debate are able to draw upon.

There are environmental impacts associated with the production of shale gas. First, the fracturing of wells requires large amounts of water which may affect the availability of water for other uses and aquatic habitats. Second, contamination of aquifers and ground water is possible, with gas emanating from taps being lit to show the effect; but methane has been seeping from the ground and such stunts as lighting spring water have been demonstrated for many years before fracking. Shale formations are typically at least one mile below the water table, trapped below many layers of impermeable rock. Fracking usually only occurs about 100m around the pipe, so faulty well construction, or other exposure pathways, are the only means by which methane could be inadvertently released. Third, if mismanaged, fracking fluid, which may contain potentially hazardous chemicals, may be released by spills and leak and such releases may contaminate surrounding areas. Fourth, fracking also produces large amounts of wastewater, which may contain dissolved chemicals and other contaminants that could require treatment before disposal or reuse. Because of the quantities of water used and the complexities inherent in treating some of the wastewater components, treatment and disposal are an important and challenging issue. Fifth, fracking may cause small earthquakes, but these are almost always too small to be a safety concern.
concern, and certainly no greater than the many earthquakes caused by traditional mining activities.

The UK’s Royal Academy of Engineering and Royal Society concluded all of these risks were very low and could be adequately addressed through proper regulation and oversight56.

Last, it is claimed that, although shale gas only releases about 50% of the CO2 compared to coal when burnt, the extraction techniques result in the release of methane; as methane is up to 70 times more harmful as a greenhouse gas than CO2, then any benefits on burning are outweighed during production. The most recent study concluded that, whilst this may have been the case with early shale gas wells, improved technology allows for greater capture of the gas thus overcoming the negative effects and shale gas is cleaner than coal overall57.

China

The most significant exploitation of shale gas would be if China was able to unlock its vast reserves. As recently as 2006, China was an exporter of natural gas, but with rapid economic development, has become a large importer. Production has risen to 3.6 tcf compared with consumption of 4.6 tcf, although natural gas still only provides 4% of its energy needs58.

China has drilled about 100 test wells, but the shale gas there is less recoverable than that in the USA and, therefore, more expensive to extract. These challenges are among the factors that have caused China to fall behind its own shale gas targets. China is relying on gas as one of the major fuels (with nuclear and hydroelectric) to supplement and supplant dirty coal in the future with demand expected to triple to over 11 tcf/y by 2035. In November 2009, USA President Barack Obama agreed to share USA shale gas technology with China, and to promote USA investment in Chinese shale gas development59. In addition to increasing domestic production, it will increase its imports of gas by LNG60 and pipeline, and will need to establish a wider natural gas network61 and storage capacity.

China has an ambitious programme of pipeline development, mostly to import natural gas from Turkmenistan, Uzbekistan, Russia, Kazakhstan and Myanmar; this entire additional pipeline infrastructure could also support China’s nascent shale gas industry. The greatest difficulty for China is that its shale gas deposits are mostly in the north-east and north-west, which are water-depleted and, therefore, large scale exploitation is unlikely until alternative water supplies become available or alternative fracking fluids are developed. The industrial environment in China is also poor, with state-run giants controlling access to services and pipelines, where they exist, and inadequate (for Western investors) land and intellectual property rights62. China is unlikely to meet its own ambitious targets for shale gas extraction, but production is still predicted to triple by 203563.

Russia

The most significant producer of shale gas, after the USA, could be Russia which holds the world’s largest reserves, with 1,680 tcf, accounting for about a quarter of the world’s total, the majority of which are located in Western Siberia. However, the Russians face a dichotomy between exploiting their own resources using modern fracking techniques, and continuing to lobby European countries and companies against fracking, in order to maintain their export markets. Russia has used this economic lever of power, principally by refusing authorisation for its state-controlled and state-influenced energy companies from dealing with Western companies that are involved in gas exploration in Eastern Europe. The most recent example of this was the withdrawal of ExxonMobil
from shale gas exploration Poland, in order for it to pursue work with Rosneft in Russia’s Far East instead.

**Other producers**

Canada and Mexico, as contiguous states to the US, have similar large, exploitable shale gas reserves. With low production costs, at about $4/mbtu, and access to the Pacific seaboard, Canadian LNG would be cheaper than East Africa’s or Australia’s and therefore sell into the expanding Asian market. A number of export LNG terminals are planned on the west coast, with investment from China, Korea and Japan. Most of Mexico’s shale gas resources are in the northeast and east-central regions of the country. The Burgos Basin is contiguous with Texas’ Eagle Ford shale deposits, which has been a prolific source of production in Texas. In South America, Argentina has potentially the largest reserves, but requires external (Western) technical and financial assistance as the current administration’s treatment of foreign companies, such as the sequestration of Repsol’s assets, means that exploitation will not start before at least a change of administration.

Although not shale gas, both East Africa and Australia have large conventional gas reserves. What is relevant is that these have become exploitable only because of the burgeoning LNG market. Each has about 200 tcf, which could produce an estimated 30 tcf/y (US annual consumption) from about 20 trains. Shell’s Prelude is the world’s largest floating structure at 600,000 tonnes deadweight, and will be the world’s first combined drilling platform and LNG train on a ship which is being manufactured to exploit gas reserves 120km off the North-West coast of Australia. Australia will overtake Qatar to become the largest LNG producer by 2018 and Africa will overtake the Middle East to become the largest net LNG exporter by 2028.

**Europe**

Europe could be one of the greatest beneficiaries of fracking for gas, but there is widespread opposition to its use in many countries, and European regulators, spurred on by MEPs who voted for new laws requiring drillers to meet stringent environmental conditions, could kill off the industry in its infancy.

Britain, Ukraine and Poland are in the vanguard of exploring the potential for shale gas. There are estimates that the UK has 1300 tcf of gas in one field alone, equal to 50 years consumption. Whilst exploitable reserves could be as low as 10% of this, shale gas could meet a significant proportion of UK demand, reduce energy bills, boost the economy of northern England and reduce carbon emissions by replacing coal and dirtier gas imports. UK shale gas extraction costs would be in the range $7 -12 mbtu: At the lower end of cost it would provide cheaper natural gas than imported, at the higher end of cost it would be comparable to European (Russian) gas, but higher than imported USA LNG. However, to match current imports of 3 tcf/y would require over 10,000 active wells with all the legal, planning, environmental and industrial implications, which is most probably infeasible.

One estimate is that shale gas could add $3trn to Europe’s GDP, create up to 1 million jobs by making industry more competitive, and reduce energy imports.

Overall, the most significant disadvantage is that Western Europe is much more densely populated than the USA, so disruptions or intrusions into daily life will require careful handling to gain public acceptance, especially if, unlike the USA, landowners do not have a direct stake in the proceeds. Europe also has a higher clay content of the shale bearing rock compared to the USA which makes...
Global implications of shale oil and gas

There are a number of trends resulting from the shale oil and gas revolutions that are already clearly apparent and set to continue. Other deductions may be made based on a realistic scenario of increased shale oil and gas production, reduced imports of crude products and increased exports of refined petroleum products, and the changes to the energy mix in the USA. These trends will largely affect the USA and Europe/Russia. The further deductions are from assuming little or no worldwide shale oil extraction, but significant and widespread exploitation of shale gas, albeit over a period of decades rather than years as happened in the USA.

Oil Price

The oil price is unlikely to be affected directly by the shale oil revolution. The impact from the revolution in the USA is already factored into the world oil price and, even though shale oil production there is large, the projected increase of up to 2 million b/d by 2020 is less than expected from either Iraq or Iran, and about the same as the spare capacity in Saudi Arabia. There have been local price effects, such as in the Gulf of Mexico where the petrol refineries and chemical companies were suffering a glut of USA shale oil that, because it is lighter than the crude oil from Saudi Arabia and Venezuela that they traditionally imported, they were struggling to process. This has led to temporarily lower prices for USA oil compared to the wider international market price. However, the USA refineries and chemical companies have started to expand in order to make better use of this oil, in order to export the refined petroleum products, and the price differential has since narrowed.

Natural gas price

The natural gas price, because it is less fungible than the oil price, has been suppressed in the USA. This has led to a divergence in price between USA, Europe and Russia, and the Far East. The benchmark USA price has been around $34/mmbtu and is expected to rise to $5-6/mmbtu as suppliers find either increasing uses for it in their refineries, or it is exported as LNG. Even at this price, it equates to only about $35 per barrel of oil (on an equivalent energy basis) and about $70 per barrel of oil once liquefaction, transport and regasification costs are included. It is this price differential that is spurring increased petro-chemical production from gas in place of oil in the USA in particular. Worldwide gas production is predicted to quadruple by 2040, with both conventional and shale gas production contributing and, more significantly, increased diversification of trading partners for both exporters and importers due to LNG. Increased fungibility of the gas market will provide greater resilience of supplies and some narrowing of the gas price differentials worldwide; however, the costs of the LNG infrastructure and transport will remain significant and thus a fully fungible market is not achievable.

European industry

European industry will be the biggest loser in the shale revolution. Europe’s refineries and petrochemical plants were already under threat from lower cost producers in Russia and the Middle East, particularly as their oil exporters have started to move up the value chain to produce refined products rather than export the raw material. Europe also has higher energy costs, partly as a result of policies to reduce CO2 emissions and subsidise renewable energy sources. Now the USA has become both a lower cost producer of petro-chemicals, due to lower feedstock and energy costs, and become an exporter of refined petroleum products and, shortly, of LNG. Most European
petro-chemical plants use naptha (oil-based) as feedstock, with only 4 out of 46 using ethane (gas-based).\textsuperscript{75} The large price differential between oil and gas puts most European plants at a disadvantage even before the higher energy costs are factored in. Grangemouth’s plant, although using North Sea gas as a feedstock, is uneconomic and only investment to allow import of LNG from the USA will keep it open\textsuperscript{76}. An estimated $90 Bn investment in the USA’s chemicals industries has been announced, equating to almost 10\% of world capacity\textsuperscript{77}; most of this increase will come at Europe’s expense.

East goes East, West goes West

Currently, oil and refined petroleum products are traded globally and gas is traded regionally. Exploitation of shale gas, together with LNG technology, could see both commodities being traded globally with effectively a single price set by cargo rather than destination. However, the costs of LNG transport for gas rise with distance, which is not the case for oil and oil products and which, therefore, makes the LNG market as much a factor of distance as the cost of extracting the raw material. With increased oil extraction in the Americas and West Africa, the Suez transit is becoming less important. Similarly increased gas extraction, both conventional and shale, together with the costs of transport, could also reduce the significance of the Suez transit for LNG\textsuperscript{78}. Middle East oil and LNG will go increasingly to the East, as will southern gas from East Africa and Australia. The Americas will be both self-sufficient and export oil and gas either from the west coasts to Asia, or from the east coasts to Europe. Europe will also be served by west and south-west African supplies, whilst Asia will also be served from Russia’s Far East for both oil and gas.

US isolationism

A new era of USA isolationism, on the back of energy self-sufficiency, has already been declared. Whilst there are clear economic benefits to the US, the global interconnected trade and financial system relies on well-oiled markets and balanced supply and demand of energy. No one country, even the USA, is able to isolate itself from the global system, therefore the USA will remain engaged in keeping trade flowing. Greater home produced supplies would, however, increase USA resilience to supply side shocks, together with its reserve of about 400 million barrels of oil which is nearly two months of total oil imports, or over 9 months of Middle East oil imports.

Reduced OPEC influence

The USA is becoming more self-sufficient in energy, and is now producing more than it imports. However, as discussed earlier, the USA will remain a net oil importer, albeit more or all of that oil could come from other North and South American countries. Both the USA and Russia will probably match or even overtake Saudi Arabia as the world’s largest oil producers this decade, thus reducing OPEC influence in the oil markets, but global demand is continuing to rise and thus it is only a relative decline. Total global demand is predicted to reach 107 million b/d by 2030, with China overtaking the USA as the largest consumer\textsuperscript{79}. Whilst the cost of production of Middle East conventional oil is still well below that of conventional off-shore oil, shale oil and oil sands, the rentier economies of most of the Gulf states require prices above $80 or risk political instability if the current social contract breaks down\textsuperscript{80}.

Reduced Russian influence

Whilst there has been speculation that shale\textsuperscript{81} and LNG\textsuperscript{82} together would undermine the Gazprom (Russian) business model and reduce Moscow’s influence in Europe, European pipeline imports are expected still to be around 2.5 times as high as imports from LNG through to 2020\textsuperscript{83}, albeit that some western European countries have negotiated discounts from contracted supplies with
Gazprom (reportedly 10%), which were historically linked to oil prices. The former Soviet states, such as Ukraine and Poland, have not benefitted. Russia is also looking East to Asia, where the market is still growing and, hence, Russia’s investment in oil and gas fields off Sakhalin Island and in the Arctic. The main Russian gas producing fields are in Western Siberia, so rerouting pipelines from European to Asian markets is just not feasible.

Technology application

The fracking technology, together with horizontal drilling, is being applied to other potential sources of natural gas. In Oman, BP is unlocking gas from sandstone which was previously considered unexploitable. Similar techniques are being considered to unlock gas from coal beds using Underground Coal Gasification. This technique relies on the ability to “gasify” coal, which is too deep and too difficult to mine conventionally, by pumping steam or oxygen though previously fractured coal seams. One estimate is that coal under the North Sea alone could produce gas equivalent to that from fracking shale onshore.

Increased energy diversification

More significantly, whilst oil demand is expected to rise due to the increased demand from Asia, this assumes the current disparity in oil and gas prices does not continue. As discussed earlier, gas is partly displacing oil in the US as a transport fuel and as feedstock for the petro-chemical industry. If this trend were replicated more globally, together with energy efficiency as demonstrated in Europe where demand is declining, then the oil price could fall. The trend that gas replaces coal for electricity generation is also evident in Europe, due to the policy to reduce emissions, and is expected to take place in China as and when gas production or imports rise. There is significant investment in Gas-to-Liquids plants, taking advantage of the disparity in prices between oil and gas, and these are predicted to rise to match the 5% of transport fuels that biofuels currently contribute.

Climate change

The main environmental argument against exploitation of shale oil and gas is that they are fossil fuels. However, growth in demand continues, particularly from the developing world, and shale oil and gas offer opportunities to countries outside the traditional producers and exporters, to either become more self-sufficient, or exporters in their own right. Thus shale could be a substitution rather than an addition to existing fossil fuels. Where shale gas, in particular, as part of the wider increase in worldwide gas production leading to higher demand and use, is as a substitute for either oil (for transport of petro-chemicals), or for coal (in electricity generation). Natural gas is a cleaner fuel than either oil or gas, in terms of CO2 emissions, and the effect may be seen most evidently in the USA, CO2 emissions are falling, are currently 5% below 1990 levels, and predicted to remain there despite rising energy consumption as the switch from oil and coal continues. A switch of just 1% of global coal consumption to gas consumption would have the same impact in reducing CO2 emissions as increasing current renewable capacity by 11%; evidently this switch would be much easier and cheaper to implement than the latter (though they are not mutually exclusive). Natural gas is thus seen by some as a bridge fuel, slowing the growth in CO2 emissions even as fossil fuel consumption rises, and providing time for investment in alternative technologies to drive their costs down.

Conclusions

The fundamental changes wrought by the shale oil and gas revolution could be profound. Already the revolution in the USA has had significant economic impact and put the USA on a realistic path
toward a 40 year goal of energy self-sufficiency. There are two shale revolutions ongoing, in oil and
gas, and a third in the transport of gas via LNG. The revolution started with increasing production
of shale oil and gas in the USA using the process of hydraulic fracturing, but is only slowly
extending to other countries despite discovery of significant reserves worldwide. There are
significant obstacles to the exploitation of oil and gas by fracking, including environmental
concerns, the requirement for large quantities of water, and the necessary industrial, economic
and legal environments.

Although USA oil production is projected to continue to rise over the rest of the decade, the USA
will remain a net oil importer, albeit North and South America together could be self-sufficient in
oil within a few years. The USA will continue to import some crude oil from the Middle East,
because the type of oil is suited to its refineries and, as a result of the ban on crude exports, the
USA will see increased exports of refined petroleum products. Although not self-sufficient, the USA
will have greater resilience to supply shocks and higher energy security. The Suez Canal will be a
less important strategic transit for energy as a result. However, even if the USA in insulated from
Middle East supply shocks, it would still feel the effects of higher oil prices and the fall-out from the
impact on the global trading and financial systems. Thus the USA, irrespective of its own oil supply
situation, would remain highly susceptible to any global supply shocks and, therefore, continue to
have a national strategic interest in the level of oil production and its secure distribution
worldwide.

The largest direct impacts on the USA from increased shale oil production are economic; an
improvement in the balance of payments; the renaissance in heavy industry and the railways in
the mid-west; and an increase in refined petroleum product exports. The last is already showing its
effects in Europe, where refineries are under threat.

There are number of significant reasons why the success of the USA in exploiting its shale oil
reserves will not be repeated in the rest of the world: Industry is not established as it was in the
USA, with far fewer capable drilling rigs; the population is not always in favour of fracking; the
water resources required for fracking are not always available; the pipelines or infrastructure to
transport the oil to the refineries are not in place; and landowners do not own the mineral rights
which has encouraged acceptance of drilling. Although there are estimated to be up to 350
billion barrels of shale oil worldwide, compared to conventional reserves of over 1,400 billion
barrels, shale oil production outside of North America will only reach 2 million b/d by 2030.
Therefore, shale oil is not going to be a significant resource, except for the USA, for at least two
decades and the strategic implications of shale oil alone are confined to the national economic
boost and the reduced significance of the Suez canal as an oil transit chokepoint. The oil price will
be largely unaffected directly by the limited shale oil revolution. As global demand increases,
OPEC will remain influential, albeit the market will be more disaggregated with more suppliers.

The shale gas revolution in the USA is, however, much greater than that in shale oil, with greater
strategic significance. USA shale gas production is forecast to increase by 50% by 2040, and the
USA is predicted to become an overall net exporter of gas within 2 years. The rapid increase in the
supply of natural gas from fracking has made USA natural gas much cheaper than anywhere else in
the world because the gas market is not very fungible. World prices vary from up to 5 times more
expensive in the Far East and 3 times in Europe. The result has been dramatic: Gas has displaced
coal in electricity generation; manufacturers and heavy power consuming industries are investing
in plant in the USA rather than Asia; vehicles are switching to gas from oil and biofuels; the
chemical industry is expected to double in size by 2020; and, because natural gas combustion can
emit less than half as much CO2 as coal combustion, USA carbon emissions have fallen in the last
few years.
LNG is a complementary aspect of the shale gas revolution as shale gas allows exploitation of previously unexploitable reserves whilst LNG provides the means for more flexible transport than fixed pipelines, thus creating a more fungible, global market for both source of supply and pricing. The prospect of USA natural gas entering the market is changing the way natural gas has been traded globally for years. Traditionally LNG prices were linked to oil prices, but now some contracts for supply are partially linked to the USA benchmark natural gas price. The advent of LNG, as a means to get gas to the market without a land-based pipeline system, will allow Australia and Africa to become major gas exporters in the next two decades and, overall, there will be increased diversification of trading partners for both exporters and importers of LNG. However, despite higher imports of LNG into Europe, Russia will remain influential as pipeline imports are expected still to be around 2.5 times higher than imports of LNG through to 2020. Ukraine and Poland will continue to rely on Russian gas, as it will take many years to exploit their own shale gas reserves.

World shale gas reserves are estimated at 50% of all other existing reserves (equivalent to 100 years supply at current global consumption rates), but whether they will be exploited depends in many countries as much on the environmental issues as it does on the logistics and economics of extraction. Fracking requires large amounts of water, contamination of aquifers and ground water is possible, fracking fluids may be spilt, wastewater requires careful disposal or treatment and fracking may cause small earthquakes. However, independent reports have concluded that all these risks are very low and could be adequately addressed through proper regulation and oversight. Worldwide shale gas production is predicted to quadruple by 2040, with the USA contributing about 50% of the total production, but only Russia, China and Argentina are expected to produce significant quantities of shale gas. Europe could be one of the greatest beneficiaries of fracking for gas, but there is widespread opposition to its use in many countries, and European regulators could kill off the industry in its infancy.

The fracking technology, together with horizontal drilling, is being applied to other potential sources of natural gas, further increasing the availability of supplies and suppliers. The gas to oil price disparity is expected to continue and, therefore, gas will displace oil as a transport fuel and as feedstock for the petro-chemical industry. Increasingly gas will replace coal for electricity generation. As gas emits only about 50% of CO2 compared to oil and coal, CO2 emissions will decline relative to where they would otherwise be, albeit that gas could also squeeze out more expensive renewable energy providers.

The USA natural gas price is expected to rise as exports increase, but world gas prices will remain at an equivalent of about $35-$70 per barrel of oil even after liquefaction, transport and regasification costs are included. European industry will be the biggest loser in the shale gas revolution, as Europe’s refineries, petro-chemical plants and energy intensive industries are uncompetitive compared to those in Russia, the Middle East and USA.

Endnotes

1 1 Barrel of oil = 42 US Gallons
3 Wells are typically drilled to a depth of several thousand feet
4 EnergyfromShale.org
5 An average well requires 2 – 4 million USA gallons (8 – 12 million litres) which is equal to the water in 3 – 6 Olympic sized swimming pools or about 200 -400 standard lorry tankers.
6 EnergyfromShale.org
7 Address given by Richard Nixon (7 November 1973) in response to the oil crisis, announcing his goal of energy independence by 1980 (Project Independence).
Shale oil is not the same as oil shale. The latter comes from a process, dating back to the 10th Century, of crushing oil bearing shale above ground, whereas shale oil is extracted by fracking underground and is sometimes referred to at tight oil. This paper is concerned only with tight oil/shale oil.

b/d is barrels of oil per day. A barrel of oil holds 40 USA gallons and is the standard volume for oil production and trade.

USA EIA Petroleum Statistics 1949 – 2011
USA EIA - Annual Energy Outlook (AEO) 2013.
BP Energy Outlook 2030 dated 2013. USA demand is expected to fall to 16.5 million b/d by 2030.
USA EIA - Annual Energy Outlook (AEO) 2013.
USA EIA US Monthly Crude Oil Imports – Dec 2011
USA EIA International Energy Outlook – Canada Dec 2012
Mexico and Venezuela produce a heavy type of crude, similar to that from Saudi Arabia, which US refineries are optimised to refine.
USA EIA - Annual Energy Outlook (AEO) 2013.
USA EIA - Annual Energy Outlook (AEO) 2013. Most of the crude oil (1.5 million b/d out of 1.7 million b/d) is transported by the SUMED (Suez- Mediterranean) pipeline.

A full rig is estimated at $30million.
USA EIA total 2009 Distribution of Wells by Production Rate Bracket. Albeit half the active wells produce less than 10 barrels of oil a day, but the nodding donkey is a recognised feature in many states.
USA EIA US Crude Oil and Natural Gas Rotary Rigs in operation 1973 – 2013. From a peak of 4000 active rotary drilling rigs in the early 1980’s, the number of active rigs dropped to only 400 by 2000, thus there was spare capacity to react to both the shale oil and gas booms.
USA EIA AEO Jun 2013 - Technically Recoverable Shale Oil and Shale Gas Resources: An Assessment of 137 Shale Formations in 41 Countries outside the United States.
BP Energy Outlook 2030 dated January 2013
USA Oil and Gas Journal 3 Sep 07 p35. A trillion cubic feet (TCF) is the standard unit for gas production and consumption at the country/international level. Normally quoted as tcf per year (tcf/y)
USA EIA AEO 2013.
USA EIA AEO Jun 2013.
ibid
Cheniere Energy’s Sabine Pass venture in Louisiana has contracts signed with UK, Spain, Korea and Japan.
FT 16 Sep 2013 – these 4 licences equate to 2.3 tcf/y about 10% US production with a further 2.8 tcf/y in application.
The price is quoted on the spot market as the Henry Hub price. Henry Hub is USA natural gas trading centre.
USA EIA AEO 2013.
ibid.
ibid. Diesel fuel is produced from gas using gas-to-liquids (GTL) technology.
ibid.
ibid.
American Chemistry Council reported in the Daily Telegraph 7 Sep 2013.
USA EIA AEO 2013
The first liquefied natural gas plant was built in West Virginia in 1912, while the first commercial liquefaction plant was built in Cleveland, Ohio, in 1941. In 1964, the British Gas Council began importing LNG from Algeria to Canvey Island, making the UK the world’s first LNG importer and Algeria its first exporter. Since then over 200 liquefaction and re-gasification trains have been built worldwide.
A typical LNG plant, called a train, costs about $3bn, with a receiving train at about $2bn. The world’s largest LNG facility at Ras Laffan, Qatar has 11 trains. Typical costs for liquefaction, transport and regasification are about $6 /mbtu (Emst and Young Oct 13).


USA EIA World Shale Gas Reserves – 2012 (EIA ARI World Shale Gas Resources)


Oil and Gas Journal 110.1a (Jan 9 2012) – Dittrick, Paula

UK’s Royal Academy of Engineering and Royal Society - Shale gas extraction in the UK: a review of hydraulic fracturing
Issued: June 2012 DES259.

David Allen, University of Texas. Financial Times 17 Sep 13.

USA EIA International Energy Outlook – China, Sep 2012


USA EIA International Energy Outlook – China, Sep 2012. China imported its first LNG shipment in the 2006, and now has 5 trains with another four under construction and more receiving government approvals.

Ibid. In order to accommodate greater gas flows from Central Asia, China will construct a third West-East Pipeline by 2015 and there are proposals for a fourth and fifth West-East pipelines.

FT 8 Jan 2014 Analysis - A new frontier by Lucy Hornby and Ed Crooks.

USA EIA International Energy Outlook – China, Sep 2012

The New Power Map – World Politics After the Boom in Unconventional Energy – Aviez Taher in Foreign Affairs, Jan 9 2013 Snapshot

Deutsche Bank estimated US LNG exports would cost $9 to $10/mbtu between 2016 and 2018, similar to its estimates for Europe’s benchmark gas. With marginal costs of around $10/mbtu Australia, the country with the highest expected export increase, also has some of the highest marginal costs in the world, so its LNG will almost exclusively go to Asian buyers.


The British Geological Survey Report 27 Jun 13

Bloomberg quoted by Daily Telegraph Feb 22 2013.

Poyry Management Consulting and Cambridge Econometrics as reported by Telegraph.co.uk on 25th Nov 2013.

Stevens, P. Shale and a Renewed Dash for Gas in the UK?. Chatham House – Expert Comment 11 Dec 2012

USA EIA AEO 2013.

BP Energy Outlook 2030 – Energy Units Conversion Tables.


Ibid.

Daily Telegraph 9 Dec 2013

Ibid. Grangemouth costs are currently about $1300 a tonne of ethylene produced against a market price of $1275. Costs based on USA LNG should be around $650 a tonne, compared to USA costs of $450 and Middle East costs of $350. The most efficient European plant is currently in Norway with costs of $950 a tonne.

Sunday Telegraph 15th December 2013.

The Suez transit took 14% of world LNG cargoes in 2012 - US EIA

BP Energy Outlook 203 dated 2013 – USA demand is expected to decline slightly to 16.5 million b/d from 18.5 million b/d currently, whilst Chinese demand increases from today’s 7 million b/d to 17 million b/d.

RCDS Lectures various RCDS 13/14.

Op Cit – BP Energy Outlook 2030. Optimistic forecasts of European shale gas extraction are for shale gas to increase total gas production by 1% per year, which does not meet the annual growth of demand of 1.4%.

Op Cit BP Energy Outlook 2030. There about 30 LNG regasification trains being built or planned in Europe, but much of the LNG being imported will either replace declining production from the North Sea or meet increased demand for electricity generation, rather than displace Russian gas.

BP 2013 Energy Outlook to 2030.

The New Power Map – World Politics After the Boom in Unconventional Energy – Aviez Taher in Foreign Affairs, Jan 9 2013 Snapshot

Daily Telegraph 16 Dec 2013.

Sunday Telegraph 15 Dec 2013.

BP Energy Outlook to 2030 dated 2013.
Is the Black Sea an area of future conflict or cooperation and what are the implications for regional maritime security?

Lt Cdr T J Davey RN

“The Black Sea region is a contested neighbourhood and the subject of intense debate. This reflects the changing dynamics of the region, its complex realities, the interests of outsiders and the region’s relations with the rest of the world. Its strategic position, linking north to south and east to west, as well as its oil, gas, transport and trade routes are all important reasons for its increasing relevance.”


Introduction

The Black Sea has a rich, diverse and volatile history shaped by the clashes of major historical powers from the Greeks, as early as 700BC, through the Roman, Byzantine, Persian and Ottoman Empires to the Soviet Union and modern day Russia. Significant geopolitical changes occurred as a result of the collapse of the Soviet Union in 1991, as many former states attempted to leave the orbit of the Kremlin, embarking on a symbolic national path to the West through closer association with NATO and the EU, to redefine the strategic landscape. Russia’s recent offer of substantially discounted prices on gas supply to Ukraine led to their president turning down an EU Association Agreement that sparked widespread civilian protests in December 2013. Just months later, Russia annexed Crimea and reclaimed the territory, undoing the gesture by Soviet leader Nikita Khrushchev, to grant Crimea to the Ukraine in a largely administrative act in 1954, marking the 300th anniversary of the Cossack army submitting Ukraine to Russian rule. The referendum determining Crimea’s Russian future in March 2014 has potentially sparked a wider proxy battle for the control of other former Soviet Union countries and, perhaps, represents the “greatest contest since the end of the Cold War” in the Black Sea region. Georgia experienced the wrath of Russia in 2008 as their dispute over South Ossetia erupted into a short but brutal war, reinforcing Russia’s dominant position over the area and necessitating an EU negotiated ceasefire to stop the fighting.

The 2014 Winter Olympics in Russia’s Black Sea coastal resort of Sochi should have been an opportunity to stimulate regional cooperation around the Black Sea and, more widely, to strengthen international dialogue with Russia to address substantial security issues in the Caucasus. Two large bomb detonations in the southern Russian city of Volgograd in the build up to the Olympics were a stark reminder of the separatist threats, most likely from Chechnya, to regional security. The Russian Foreign Ministry was prompted to call for “international solidarity in the fight against an insidious enemy that can only be defeated together”. As events have transpired, in particular over the Crimean peninsula, international cooperation is disintegrating. Russia has been suspended from the G8 forum and the US and EU have imposed strict trade and economic sanctions against Russia, targeting business, military and government officials for asset freezes and global travel restrictions. In this context, this paper sets out to explore whether the Black Sea is an area of future conflict or cooperation, with particular reference to the implications for maritime security in the region. The Commission on the Black Sea, quoted above, has reported how the region is complicated, dynamic and strategically situated to ensure it is increasingly relevant for continued study.

The paper is structured in four chapters to build an understanding of the Black Sea region and develop some observations for maritime security in the future. In the first chapter, the question of how to conceptualise the Black Sea region is addressed. Due the existence of competing definitions, structures and organisations which all add to the complexity, the answer is not commonly agreed. Making
reference to both historical and contemporary analysis, the aim is to provide a Black Sea region definition that is valid for analysis in the paper, whilst addressing external and global pressures on the region.

The threats to stability and conflict in the region are addressed in the second chapter. With an overview of the frozen conflicts, potential spill over from Caucasus unrest, and the threat of extremism; there is much to unsettle the region. The international crisis over the Crimea is not analysed in depth due to its rapidly changing nature at the time of writing, but is acknowledged as a potential catalyst for further instability.

The third chapter examines the current cooperation initiatives and focuses, in detail, on the Black Sea Economic Cooperation (BSEC). A wide array of organisations exist throughout the region, all aimed at building cooperation in specific sectors or disciplines. It is outside the scope of this paper to examine them all, but some small case studies will demonstrate that progress can be made, if the right teams are engaged, and a national emphasis is placed on leadership, decision making and action. The framework, membership, aims and outputs of BSEC, since its inception, provide an example of partial success and are indicative of cooperation prospects in the future.

The final chapter addresses maritime security in the Black Sea. First, the concept is defined and risks to security at sea are identified. Then the wider implications for cooperation or conflict at sea are investigated, particularly through analysis of the Black Sea Force Naval Task Group (BlackSeaFor), the Operation Black Sea Harmony (OBSH) initiative and some of the other regional programmes that attempt to address the principal challenges at sea.

Research has been conducted primarily through an extensive literature review to gain an appreciation of the Black Sea as a region and its wider influences. Throughout, a balance of argument has been sought and while much of the literature is of Western origin, many of the authors have personal experiences of living or working in the region. In conclusion, the paper will argue that, whilst conflict is the current prime feature of the Black Sea, renewed effort must be made to re-introduce meaningful cooperation to safeguard a prosperous future for the region. Notwithstanding the Crimea crisis, it is possible to predict a future where the stakeholders work in unison to maximise the opportunities at all levels from geopolitics to local businesses. Maritime security responses provide a useful tool to analyse where and how this process can begin. The Black Sea Commission agrees in principle;

“we assume that concerned actors are willing to explore ‘win-win’ options that permit the realisation of mutual gains and are not locked into ‘zero sum’ or relativist ways of thinking, in which one party’s gain is automatically perceived as another’s absolute or relative loss.”

Cooperation can become the norm. For example, Henry Kissinger has recently suggested that Ukraine can only be helped if Russia, the EU and the US all work towards agreed outcomes and not resort to competition over political posturing. His opinion chimes with a pro-Russian government journalist, who notes how his proposed tri-lateral actions should lead to future optimism, an optimistic might “drag the Ukrainian economy and statehood out of the dire straits into which they have fallen”. In contrast, others argue that many of the Cold War scenarios will make a comeback and a prolonged period of ‘armed peace’ will follow, with Russia becoming “a nuisance power, blocking Western initiatives”. This paper disagrees, particularly when focused solely on the Black Sea and unlocking the potential it holds.

**The Black Sea region**

“On the shores of the Black Sea, there were born a pair of Siamese twins called ‘civilisation’ and ‘barbarism’.”

Charles King describes the 135 years since the end of the Crimean War, in 1856, to the collapse of the Soviet Union as “a time when politicians and planners laboured to unmake the Black Sea as a region”. His thesis is broadly correct and reinforced by other commentators. Aybak contends that the study of the Black sea as a region is a “largely neglected phenomenon”; in 2010, Khara similarly argues that the
region has been overlooked or only used as “a tool for trans-regional purposes”16 and, most recently, Weaver writes that the ‘sea’ is at the “centre of a region of nations which are learning to share it”.17 As the Black Sea now tends towards civilisation rather than barbarism, notwithstanding Russia’s action over the Crimea, understanding how and why the region operates as it does, supports the subsequent analysis.

Regionalism is defined as the “development of institutionalized cooperation among states and other actors on the basis of regional contiguity as a feature of the international system”.18 Regions themselves are often nebulous in their construct and also regularly change over history, although cultural, linguistic and historical ties tend to dominate why states join together. The processes underpinning regionalism and interdependence generally include a mix of economic, social, political and security concerns.19 There is also an important distinction between cooperation and integration. Cooperation represents limited arrangements agreed between states in particular fields, such as energy, transport or combating crime.20 Integration refers to the formal processes which dismantle barriers for interaction, create rules and promote joint governance.21 The latter tends to materialise through economic unification, such as in free trade areas, customs unions or a common market, and is ultimately characterised by political harmonization. The EU exemplifies advanced regional integration with monetary union, widespread regional governance and a strong legal framework, giving it a unique status as an international actor in its own right.22 Conversely, the Black Sea region, at best, only reflects cooperative schemes and, as the paper will demonstrate, progress has been relatively slow to mould the members into a cohesive and functioning regional entity.

An historical perspective provides some context to understand the complexity of the region. The Greeks first set up trading posts on the northern Black Sea coast in the eighth century BC,23 marking the start of Ascherson’s gestation period for the twins he identified. As the emerging Greek civilisation encountered northern shore indigenous populations, it marked the first phase of wider colonisation and began the process of ‘Western Civilisation’ of the region.24 Hitchner identifies a further four phases of Black Sea history. First, the Roman Empire and its sprawling globalisation process which brought security through military presence, introduced the consumer based economy, shaped provinces and effectively transformed the ancient world into a civilian society.25 Second, the regional transformation triggered by the foundation of Constantinople on the Bosphorus and the dominance of Byzantine economic expansion in the region. The legacy of the Black Sea as a “great north-eastern commercial and cultural frontier of Europe”26 has its roots in Byzantium. The transformation continued through with the spread and dominance of the Ottoman Empire, marked by their capture of Constantinople in 1453 and the start of 300 years of Ottoman hegemony, before the sea itself was opened up to European commerce in the late eighteenth century.27 The emergence of the Czarist Russian Empire, a declining Ottoman Empire and competition with the West, culminating in the Crimean War from 1853 to 1856, signify the final phase. The Black Sea remained the focal point as the balance of power shifted north to Russia and then the Soviet Union setting the tone for the remainder of history until today. The last stage was a period of developing ethnic identities, changes in ideological influences and, ultimately, the development of a dysfunctional model in terms of economic structures.28 Constructive regional achievements over time, and a tendency towards prosperity, appear to have been undone in the final phase of Black Sea regional development.

Özveren suggests a harsher prognosis and argues that, despite some of its superficial successes, the Soviet Union introduced a large economic burden to the region. As such, the “Black Sea remained fractured, paralysed and defunct as a zone of capital accumulation until the dismemberment of the Soviet Empire”.29 Some of the newly emerging states, especially those with a Black Sea coastline, sought to exert influence in their maritime domains. However, the kingdoms of Romania and Bulgaria; the briefly independent republics of Georgia and Ukraine; and Turkey as a republican successor of the Ottomans, were prevented from forming any sort of regional alliance or developing much needed cohesion due to the dominance of Soviet Communism.30 In effect, the sea became the front line of the global conflict between NATO and Western capitalism, and the Warsaw Pact and communism.

More recently, the opportunities offered by the collapse of the Soviet Union have yet to be fully realised. The new Black Sea countries are effectively modern states that have been competing in state building
projects, developing their industrial concerns, maximising the available resources and exploiting the sea to build large population centres and port cities. However, the ability of those states in the last 23 years to cooperate has been limited and proved a major challenge. As such, “economic crisis, state weakness and military conflict have come to characterise the broad region of the Black Sea basin”.

It follows that a shared sense of the diverse and rich history must be retained but not dominate the future of Black Sea politics. Linked to this, King has researched the development, perhaps before its time, of a think-tank style organisation prominent in the mid-1920s called the Promethean project. It proposed an alliance of the Black Sea states (then Bulgaria, Romania and Turkey) to counter the regional hegemony of the Soviet Union, and also campaigned across southeast Europe in protest over the incorporation of Ukraine and the Caucasus states into the Soviet sphere. The project eventually failed as support dwindled, primarily due the events of the Second World War, and by 1951 it ceased to exist in any form. However, the founding principles that brought the retired politicians and diplomats together to issue the first of their *Prométhéé* journals in 1929 have resonance today. The notion that the “Black Sea community would become little more than a quixotic project of émigré cold warriors”, can be partially discounted today as a new sense of regional identity emerges.

The academic literature on the Black Sea region does not agree on how to define to the region itself. Manoli views it as a ‘sub-region’ and one to be considered through the lens of a wider European integration perspective, mainly focusing on the twelve BSEC member states but suggesting that the EU influence is predominant. Henderson suggests that the six Black Sea littoral states form the crux of the region and then adds Armenia, Azerbaijan and Moldova to encompass a ‘wider’ Black Sea region. Others broaden the definition, offering that, from a Turkish perspective, the wider Black Sea region is one that connects the Caspian, Aegean and Mediterranean Seas, or when viewed from Washington, a vast region that also encompasses the Caspian and Baltic Seas. Romanian commentators have more recently defined the region in the context of BSEC but add that it is not about delineating a territorial expanse, instead suggesting that it is a “grey area...[representing]...the geopolitical frontier between Russia and the Western world”. This theme follows in Triantaphyllou’s analysis; geopolitical trends are critical and have influenced the BSEC region from within, dominated by Turkey and Russia as the regional hegemons. His argument builds on Buzan’s theorising on regional security complexes and necessitates a view that the Black Sea region is defined by the intricacies of the member’s interlinked security concerns. To simplify the study, this paper acknowledges that the region’s classification is contested and can only really be defined by the strength of the relationships that exist within it, rather than the identity of its constituent parts. To examine cooperation and security, BSEC membership will be utilised and for more detailed maritime security analysis, the lens will be narrowed to just the littoral countries that can exert a direct influence on the sea.

It is agreed that the Black Sea matters and retains its strategic significance however it is defined. The region connects Europe with Central Asia and links both to the Middle East. It was part of the historic Silk Route from the Far East, and is developing into an energy hub for the distribution of gas and oil, primarily from the Caspian Sea but also from its own resources. The European expansion project puts the EU’s eastern border directly on the Black Sea’s shores through Romania and Bulgaria. The EU’s wider developmental policies in the region including the European Neighbourhood Policy (ENP), Eastern Partnership (EaP) and Black Sea Synergy (BSS) initiatives, all have a part to play. NATO’s sphere of influence is arguably larger with half of the littoral countries under its security umbrella and Georgia remains an enthusiastic aspirant nation. The EU and NATO’s strategic relationship is strong and their collective influence pervades the region. Both have strategic partnerships with Russia; NATO through the NATO-Russia Council, although at the time of writing, activity is suspended until at least June 2014, and the EU through their Partnership and Cooperation Agreements. To provide some counter-balance, Russia has sought to develop a Eurasian Customs Union and bolster their regional security through the Collective Security Treaty Organisation. Membership of both is primarily derived from Belarus and the central Asian ex-Soviet states, and their influence tends to be centred there and further east, without specifically impacting on the Black Sea region. The latter’s significance is therefore not in dispute and whether its future is sustainable and secure is examined below.
Black Sea conflicts

“Empires never collapse without leaving in their wake divided populations, disputed borders, and decades of simmering grievances”.45

This chapter examines the threats to stability in the Black Sea region and assesses their causes, impact and likelihood for resolution in the future. There are four ‘frozen conflicts’ in the region; in Nagorno-Karabakh between Armenia and Azerbaijan; in Moldova (Transnistria); and in Georgia (Abkhazia and South Ossetia). They are frozen in that, whilst their origins stem from the break up of the Soviet Union and each has been temporarily resolved, the root causes remain a disruptive factor and have the potential to resurface. This section provides an overview of each and addresses how the tensions affect regional stability, before turning to examine cooperation in the next chapter. The current situation in Ukraine is not assessed in detail due to its contemporary and rapidly changing nature, making longer term analysis of regional implications redundant.

The frozen conflicts are broadly related through themes such as a struggle for independence, the right to national self-determination, security dilemmas, ethnic divides, secessionist movements and organised crime.46 Asmus summarises their collective impacts across the region;

“They are festering wounds that absorb political energy and drain economic resources from countries that are already weak and poor. They inhibit the process of state building as well as the development of a more democratic society. They generate corruption and…are a major source of instability within these countries and in the region more broadly.”47

It follows that the conflicts need to be addressed and yet the efforts, or lack of, from a Euro-Atlantic perspective, from Russia and from a variety of international organisations, including BSEC and the Organisation for Security and Cooperation in Europe (OSCE) have not been successful. In general, a reinvigoration of the mediation process is required in all of the conflict areas. This should lead to a more prominent role for the EU or the UN, which are effectively the only institutions with the multi-disciplinary skills to revitalise the processes. It also implies that Russia would have to be further engaged and, in particular, persuaded that her own long term influences and wider regional interests are not served by the “perpetuation of rogue mini-regimes”.48 Also, the conflict areas are not necessarily ‘frozen’, a potential misnomer, as skirmishes and fighting continue on some disputed borders across the region.49

As the Transcaucasian Federative Republic dissolved in 1936, the regions of the South Caucasus became Soviet Socialist Republics and Nagorno-Karabakh was incorporated as an autonomous oblast within Azerbaijan.50 Armenia, Azerbaijan’s western neighbour, made numerous requests for the oblast to be transferred to them, claiming ownership through their demographic ties and geography, despite the territory being part of Azerbaijan. In the intervening period before conflict broke out, violence was commonplace and Soviet led negotiations stalled, until actions by Azeri mob groups against Armenians in a small provincial towns sparked an “irreparable meltdown in ethnic relations”.51 In response, this sparked vast movements of the populations across borders, claims of genocide and further ethnic clashes. Stalemate ensued until a full scale war broke out in 1992 when Armenian forces went on the offensive. This was arguably an inevitable outcome due to piecemeal efforts by both countries, Russia’s President Yeltsin and the inaction of the OSCE Minsk Group (formed to try and negotiate a peace deal).52 Despite a Russian brokered ceasefire in 1994 and UN Resolutions demanding an Armenian withdrawal, the situation effectively remains deadlocked today and peace talks continue through the Minsk Group, with close support from France and the US.53

In terms of wider regional stability, the frozen conflict of Nagorno-Karabakh has some significant implications. The structural weaknesses of the politics in the region were the main contributory factor, simply conceptualised as being a “tiny knot at the centre of a big international security tangle”.54 The cause can be attributed to the fact that the true protagonists, the people, were driven to “act by passionately held ideas about history, identity and rights… (and those)... ideas expanded inside the ideological vacuum created by the end of the Soviet Union”.55 To achieve peace, a real shift is required in
national perspectives from both sides and, to a certain extent, the people themselves. Intransigence and irreconcilable positions are likely to remain the defining feature of the conflict. A solution requires reinvigoration of a mediation process, most likely in various guises, from the Minsk Group, through to BSEC and the EU. An incremental, step-by-step process must be engendered to break down the cynicism that exists on both sides towards compromise and avoid a clash of the regional powers, in particular Russia, Turkey and Iran, that have vested, but potentially conflicting, interests in both Armenia and Azerbaijan. Regardless, “protracted conflict has high costs for both countries, including hundreds of thousands of displaced persons, bereaved relatives and closed borders” and a solution is not on the horizon.

The breakaway Moldovan republic of Transnistria is a small strip of land nestled between Ukraine’s western border and Moldova, along the Dniester River. The origins of the conflict date back to a short civil war between March and June 1992, as Moldova and the self-styled government of Transnistria led their troops into a war over territorial claims and political ideology. Transnistria favoured staying with the Soviet Union during the period of its dissolution as Moldova, with close links to Romania, secured its own independence and was developing a Western orientation. The aftermath has seen over twenty-two years of peacekeeping efforts, originally through a joint group of Russian, Transnistrian and Moldovan forces and, more recently, supported by OSCE initiatives with Ukraine as a mediator and since 2005, border management support through the EU’s Mission team to Moldova and Ukraine. Whilst the ceasefire agreement has been largely effective, it is backed by Russian troops in support of the breakaway area which is a clear point of friction, especially in relation to Ukraine.

Until the recent events in Crimea, the Transnistria conflict was arguably the least likely to thaw to a stage where fighting would resume. It is not founded in fractious ethnic tensions and has developed a limited form of stability through peacekeeping initiatives. The region is not internationally recognised, and the entrenched leadership on both sides continues to push for autonomy and boycott the negotiation rounds to avoid finding a long term resolution. Lately there have been growing, but unwarranted, fears that Russia would continue her expansionist policy and turn towards Transnistria. Transnistrians, when polled in 2006, were overwhelmingly in favour of independence and developing closer ties with Russia. Russia’s Foreign Minister has stated that the process must be led by the OSCE mediation group, who continue to meet regularly, and re-affirmed his intent that “the goal is to achieve the objective we all share – ensuring a special status of Transnistria in a united sovereign and neutral Moldova”. Media reporting and NATO’s perspective suggests that it remains a potential flashpoint, particularly if Russia continues to subsidise the area, chiefly through cheap gas supply. It will, therefore, remain relatively prosperous and Moldova will need to continue to foster her relationships with Romania and the EU, building on the current Association Agreement. Regardless of how the situation unfolds over time, whether Transnistria is subsumed like Crimea or negotiations continue indefinitely, it remains a source of instability and may yet re-emerge as a flash point.

In Georgia, the undertones of demands for independence and beginning of civil unrest can be traced back to as early as 1987, with the formation of semi-official political organisations to counter the Communist party leadership. Both South Ossetia and Abkhazia were entirely at odds with the national strategy of Georgia, favouring their own independence and closer relations with the Soviet Union and then Russia after its dissolution. Through a series of national referendums, which resulted in Georgia declaring its independence in April 1991, both regions were in turmoil having boycotted the national plebiscites. This was part of the cause of the intermittent but bloody periods of civil war at the end of the year and, by early 1992, South Ossetia had voted internally for reunification with North Ossetia (and the Russian Federation) and Abkhazia had re-adopted its 1925 constitution which re-affirmed their status as a sovereign republic. At the same time, the newly formed country of Georgia was grappling with state-building, instituting democracy and presidential elections. Despite negotiations, Georgia and Abkhazia were effectively at war from August 1992 until a Russian mediated peace deal in September 1993. It was a brutal and scarring conflict with atrocities committed by both sides, and Russia covertly supporting Abkhazian fighters throughout fuelled the discord.
Abkhazia remains unrecognised by the international community today and, whilst its physical isolation has been reduced by the re-opening of rail and transport links to Russia and the growth of primarily Russian tourism, its future is not clear. Georgia insists it is an integral part of their country and Abkhazia continues, with Russian military and political support, to make demands for sovereignty. Russia has nurtured a close relationship with Abkhazia, which now operates with its own political system under an elected president, including maintaining military bases in the area, troops on the Georgian border and by issuing passports to the Abkhazians. As Jones has summarised, a conflict that began over “constitutional issues, rights of self-government, and territorial status”; has eventually become self-sustaining, less focused on sovereignty and more about insecurity and righting perceived injustices. He contends that the powerful influence of Moscow has certainly inflated the situation to this day.

The consequences of Georgia’s war with Russia over South Ossetia in 2008 are still felt years later and will likely persist well into the future. This author’s recent visit to the disputed border, accompanying members of the civilian EU’s Monitoring Mission (EUMM) which was established at the end of the five day conflict, left very little room for optimism. Armed Russian and Georgian soldiers stand daily guard, manning checkpoints to ensure control over their parts of the territory, and small villages have been divided by impromptu fences erected by Russian troops - a so called ‘borderisation’ process. The ongoing Geneva dialogue with multilateral meetings that deal with security, humanitarian issues and mediation continues to work towards a resolution but without tangible results. As such, the situation in South Ossetia is an example of a frozen but ‘live’ conflict, although at least at face value it is a relatively peaceful one at the moment, despite the close proximity of the protagonist’s armed forces.

In 2008, this was certainly not the case. Whilst the actual cause of the war is disputed, the International Crisis Group report published shortly after the cessation, summarised that large-scale violence broke out due to “a disastrous miscalculation by a Georgian leadership that was impatient with … a Russian dominated negotiations process”. The report also goes on to criticise Russia’s disproportinate response and suggests a dramatic shift in East-West relations as a consequence. The speed and ferocity of Russia’s response and Georgia’s rapid defeat confirmed Russia’s supremacy and appetite to maintain the imbalance in Georgia, underlining Putin’s leadership and resolve. The counter offensive was not just aimed at Georgia, but also its main ally, the US; “the invading forces demonstrated their power to rampage through the country at will and took especial delight in destroying any United States-built military infrastructure”. As quickly as it had begun and five days after the first engagements, the EU’s acting President Sarkozy, led negotiations on a peace plan that brought the conflict to an end, including convening the Geneva talks and the establishing the EUMM.

As already suggested, the legacy of the 2008 war combined with ongoing tensions surrounding the frozen conflicts has far reaching consequences for the security of the Black Sea region. Jones’ summary of the situation in Georgia paints a picture that can be translated across all of the conflict areas analysed above. He argues that the war

“illustrated the historical entanglements of interethnic relations, the vulnerabilities along Georgia’s borders… and the rivalries of regional and great powers. It bared the connections between war and the economy, not just in terms of the costs, but also the links between conflict, corruption and crime. It demonstrated the importance of democratic institutions and … reinforced Georgia’s Western orientation but exposed the illusion of the West’s dependability”.

The conflicts have all been focused on vulnerabilities and rivalries, pitting the regional hegemony of Russia against the small ex-Soviet states or regions trying to establish some form of independence, by shifting away from the influence of Moscow. This almost inevitably means a more Western orientation, either in terms of governance and democracy or moves towards the gravitational pull of the EU. The economy plays a key part as well as aspects of critical energy infrastructure, particularly through the Caucasus and its linkages with the energy rich Caspian Sea. Longer term prospects are not encouraging at the national, regional and local levels. The paradox is that, through stronger democratic institutions, better political dialogue and some inclination towards compromise, years of conflict could have been avoided.
In providing an evaluation of some of the Black Sea conflicts, it is worth broadening the analysis to include wider regional security issues. Griffin identifies the main theme of the Caucasus region as one of conflict, citing Russia’s problems with its nominally autonomous republics in the south, such as Dagestan, Ingushetia and through wars with Chechnya in 1994 and 1999. He labels the region as one of the world’s most constant trouble spots and emphasises the strength of maintaining a historical perspective when assessing the conflicts. Turkey, a NATO member since 1950, is often viewed as a bridge across some significant divides; not least geographically from the East to the West, but also ideologically between Christianity and Islam. However, the analogy does not necessarily always fit especially when the divide is too wide, and Turkey has its own share of internal and near neighbourhood security issues. Cyprus, a product of the Turkish War of independence against the Greeks, remains ‘live’ today. Relations with Armenia are strained due to a disputed Armenian genocide in the First World War and normalization of that relationship appears unlikely, and may worsen in the future. Similarly, Turkey’s Kurdish populace disputes have been referred to as one of the principal challenges facing the country, again without an obvious short term solution. By extending the security lens even further to the wider region there is much more at stake. Instability, civil war and a refugee crisis in Syria, the impacts of a potential resurgent Taliban in Afghanistan, Iran’s relationship with the West and the Israel-Palestine contest could all spill over to the Black Sea region, or at least inspire more unrest and conflict. For example Bulgaria, a NATO and EU member, suffered a terrorist bomb attack in the Black Sea coastal resort of Burgas in 2011 when Israeli tourists were targeted.

This chapter has sought to demonstrate and explain how the Black Sea region is afflicted by conflict. The domination of the historical empires through to the hegemony of Russia today continues to script the narrative and fuel embittered national relationships. There is an identifiable trend in the analysis to suggest that any resolution must be viewed as a very long term objective. The complexity of the disputes, geography, culture, ethnicity and multidimensional international relationships within and across the region, all contribute to a greater or lesser extent. Crimea today is a contemporary example. Writing in 1997, Reid notes how “Potemkin called Crimea the ‘wart on Russia’s nose’, and it still itches and she speculated that if a civil war were to break it then it would be in Crimea. With discerning insight, she offered two futures for Ukraine;

“at worst, it will be a fragile state, poverty-stricken buffer-state in a new divide between and introverted West and an aggressive, unstable Russia. At best it will be a rich, heavyweight democracy in a continent-wide partnership of friendly like-minded states.”

The worst case scenario is being played out. On the tenth anniversary of his country joining NATO, Romania’s president recently stated that the frozen conflicts will test NATO’s capacity not only to preserve stability and security in the region, but also its ability to ensure Europe’s wider energy security. He also noted how the current strategic context, “marked by asymmetric trans-boundary threats, regional instability, a complex relation[ship] with Russia [and] violations of international law” add to that challenge. As such, an ambiguous approach or simply avoiding the ‘grey areas’ will not suffice. Cooperation is required and the next chapter addresses that theme.

Black Sea cooperation

“For the first time ever, the coasts of the Black Sea were to be brought together not by conquest or the informal networks of commerce, but rather by the purposeful effort of political leaders to craft a secure and cooperative region with the sea at its centre”.

This chapter begins by summarising some of the obstacles to national collaboration in the region and draws from chapter one to understand how regionalism has shaped cooperation. In that context, it examines three of the most prominent cooperation initiatives; BSEC, the (Georgia, Ukraine, Azerbaijan, Moldova) Organisation for Democracy and Economic Development (GUAM) and the EU’s primary mechanisms for interaction in the region, the ENP, EaP and BSS. A recent ‘2020 Vision’ report for the Black Sea lists 33 regional organisations and initiatives in operation, taking the broadest possible view.
Manoli identified 24 regional cooperation schemes and has also acknowledged the ‘micro-regionalism’ (as opposed to regional and sub-regional) initiatives, those “aim to enhance governance at the sub-state level and promote political and economic independence”. The existence of micro-regional schemes, such as the Black Sea Capitals’ Governors and Mayors Association and the Black and Azov Seas Ports Association, highlights how real achievements are possible at lower levels as a means to bypassing the obstacles outlined below. However, the sheer numbers of organisations are indicative of the wider regional issues and BSEC’s inability to act as the lead regional association. Complexity, duplication and inefficiencies then, unfortunately, start to become the prevailing features of the cooperation sphere.

Many of the obstacles to cooperation are due to the Black Sea being a politically constructed region shaped by risks and threats rather than purely by its core identity. The missing regional identity can be found in the crossroads character of the Black Sea area, its widespread cultural and linguistic differences, post-Soviet tensions and supra-regional competition. The latter point refers to the legacy of the Cold War front line scenario, a dynamic which has changed significantly, although the argument could equally be applied to across Western Europe, notwithstanding the less direct influences of the Soviet Union collapse. In Europe, the EU has debatably grown to become a global model for managing regionalism and promoting effective institutionalism. A significant and divergent variable is time. Black Sea cooperation began in earnest in 1992 whilst the EU’s roots can be traced to the creation of the Council of Europe in 1949 and the Treaty of Paris in 1951 to develop the European Coal and Steel Community.

Viewed from a different perspective, Black Sea cooperation is restrained by the region’s cultural, historical and geographical diversity, competing national socio-economic circumstances and the conflicting policies of state stakeholders. These factors, coupled with the erosion of political commitment, the rise or forcefulness of Russia and doubts over BSEC’s tangible unity, may point towards a bleak future for the region. However, when framed in the two distinct waves of regionalism that have occurred, the future begins to appear somewhat brighter. Manoli identifies that, whilst there is no single paradigm for Black Sea cooperation, the first wave from 1990 featured local power battles as opportunities to assert leadership materialised, the decline of the Soviet Union and the rhetoric of regionalism. In the second wave, beginning in 2000, the enlargement agenda of both the EU and NATO drove systemic political changes and the regional impact of global issues, such as terrorism, organised crime and the environment, meant that any cooperation was forced to be more inward looking. Throughout, the balance and power play between the Euro-Atlantic institutions and the remnants of Russia’s hegemony have fuelled complexity and ensured an unsteady path towards cooperation. At the time of writing, the crisis in Ukraine and a resurgent antagonistic regional power in Russia, indicate that many of the previous options for cooperation will have to be reassessed and, most likely, re-established when the dispute is eventually settled.

In 1990, the Turkish ambassador to the US announced a scheme designed to steer a path towards peace and cooperation in the Black Sea region through the formal establishment of an institution among the riparian countries. BSEC grew from here, with a vision to act:

“as a unique and promising model of multilateral political and economic initiative aimed at fostering interaction and harmony among the Member States, as well as to ensure peace, stability and prosperity encouraging friendly and good-neighbourly relations in the Black Sea region.”

The mission is commendable but significantly challenged by recent history and dynamic current events. With a wider perspective and short review of its latest activity, there remains a creditable drive towards achieving BSEC goals. For example, in March and April 2014, a snapshot of BSEC activity encompasses meetings between the Secretary General of the BSEC Permanent International Secretariat and EU Commissioner for Maritime Affairs and Fisheries in Brussels; a Communication Officers seminar in Turkey to advance the BSEC Public Communication Strategy; member state’s Ministers of Transport meeting to develop a regional transport cooperation plan and encourage sustainable economic growth through trade links; a four day diplomacy seminar to examine EU and NATO security policy for young diplomats; and a high level meeting between BSEC senior officials and the Black Sea Trade and Development Bank...
(BSTDB) directors to forge a joint economic agenda for the region out to 2020. In any analysis, this constitutes an expansive programme of events that can deliver regional benefits and should be heralded.

It is striking that there is not a single reference or statement from BSEC about the Ukraine issues on their website at the time of writing. This can be dissected from two contrasting perspectives. First, given the ongoing activity outlined above, the organisation effectively is continuing with business as usual and continues to make progress towards departmental aims. Alternatively, it could be viewed as reinforcing the restricted influence of BSEC and a sense that political officials are aware of the limitations and, therefore, should not or “cannot effectively tackle the problems of the region”. This should also be caveated with an observation that it is not intended to be a regional security organisation and therefore should not be judged solely by those parameters.

Before making a final assessment, it is worth understanding how BSEC is organised and what it sets out to achieve. The Charter, adopted in May 1999, articulates that sustained economic growth for member states has primacy, and emphasises cooperation in wide ranging and diverse areas from banking and finances to energy, healthcare, tourism, science and technology, environmental protection and combatting organised crime or cross border illicit trafficking. It has a Permanent International Secretariat, established in Istanbul in 1994, which provides the backbone of the organisation. Its focus is to ensure cooperation with other regional and international organisations; strengthen BSEC itself by promoting trade, economic development and links with industry; and consolidating the project-oriented work, notably in education, science and culture. There are also various other affiliated institutions or bodies that form the wider BSEC community. These include the Parliamentary Assembly of BSEC (PABSEC), the BSTDB, the International Centre for Black Sea Studies (ICBSS) and the Black Sea Business Council. Together they begin to appear, on paper at least, as a well organised body that could make a difference across a wide range of regional issues, effectively evolving just like the EU in its earlier guises. However, the PABSEC only meets twice a year and uses committees to provide investigative reports to feed its meeting cycle. As such, and by design, despite the protestations of some of the PABSEC representatives, its ability to act as the more established European Parliament is limited due to a Charter commitment that any mutual support is only provided on a consultative basis. It therefore lacks the teeth to deliver formal legislative decision making nor can it really impact on the political and economic issues of its members. This could work in its favour, particularly as Russia had been participating in many of the cooperative measures, when compared to her participation in EU initiated processes. However, there is a broad perspective that, due to the prevailing rivalries between members, there is arguably not a real drive to be a truly collaborative region.

As BSEC constitutes an “arrangement of heterogeneous states that are not equal in size, power, state formation, population, natural resources or level of development”, it is not necessarily going to be the most powerful or dynamic of organisations. Critics suggests that, due to significant changes in the international conditions that led to its creation, the problems of diverse transitional economies and the threats to some member’s territorial integrity, “BSEC will remain merely a dream of economic cooperation in the Black Sea basin”. Romania and Bulgaria, the western Black Sea nations, are now members of the EU and have continued to develop their ambitions for wider European integration, perhaps to the detriment of their BSEC commitment although it has clearly brought the EU physically closer to the Black Sea. Manoli assesses that some of the institutionalisation motivated purely by practical political concerns, for instance, a lack of national resources or limited administrative expertise in newly formed governments, has met some of the demands of multilateral cooperation between members. This is positive and the lower level aspects of BSEC’s achievements should be promoted more widely. Other writers have proposed options for increased democracy in BSEC to solve some of the failings. One example suggests an overhaul of the governance model of PABSEC to incorporate democratically elected regional parliament members. This would arguably start a process where regional representation could include speakers from the frozen conflict areas, which could begin to unlock some of the issues that plague cooperation and start the steps towards a functioning, practical appearance of a security community. Other initiatives have been suggested including opening up public discussion on the issues, increasing PABSEC meeting regularity and funding, open press schemes to encourage unbiased
reporting and targeting the younger generation to energise the processes. This is not without it challenges and requires “creative thinking without the creation of yet more institutions”.118

An assessment, in 2004, that BSEC will achieve its goals with the strong political will of its members119 has not materialised in the last decade. Despite much progress there are clear concerns over cohesion and achievements. It serves as a meeting place or forum for countries that may otherwise only engage in bi-lateral relations or, indeed, none at all and attempts to build relationships and grow areas of common interest where the member states can agree, particularly at lower levels. This is effectively characterised by “intergovernmental entrepreneurship rather the demands of transnational interests”120 and will most likely remain the case in the future, as any genuine political cooperation appears a distant aspiration. Somewhat paradoxically, in October 2013, Ukraine's Deputy Chairman of PABSEC and head of her country's BSEC delegation, Olena Netetska, outlined how a series of new ideas for collaboration that develop the entrepreneurial theme. How Ukraine may protest about Russia’s continued membership is not yet clear, but Netetska’s objectives make sense. Despite many of the complexities, she reinforces Ukraine’s commitment to BSEC, in particular to the project orientated activities that do have a positive impact.121 Her proposals in the transport sector include connecting all of the Black Sea port cities through maritime ferry services and improving joint railway projects. This links closely with tourism which is vital for boosting Black Sea economies, and benefits “joint promotion of the Black Sea as a single tourism brand and as a tourism macro-region in the global market”.122 Her strategy out to 2025 also includes bypassing some of the bureaucracy of the member states’ capitals, none of which are geographically near to the coastline, by fostering a Municipal Assembly of the Black Sea cities to strengthen BSEC overall. Regardless, and at the risk of promoting a new initiative, a long term perspective and coherent strategy is required, even if the Ukrainian vision survives, to see it to fruition.

BSEC tends to dominate any discussion of Black Sea regional cooperation initiatives but others do exist outside of its umbrella of structures and programmes. GUAM was created in 1997 and briefly expanded to become GUUAM when Uzbekistan was a member from 1999 to 2005, although they left due to a perceived lack of effective cooperation between states.123 With its Charter signed in Kiev in May 2006, GUAM became a formal regional organisation. The four members have a common bond and, whilst the overarching mission is noted as promoting democracy, strengthening regional stability and facilitating an energy corridor that stretches from the Caspian Sea to Europe,124 the group was also designed to provide mutual support over their territorial integrity challenges and relations with Russia. GUAM has a Permanent Secretariat based in Kiev and a structure not dissimilar to that of BSEC, just on a smaller scale. A Business Council was created in 2002 and Parliamentary Assembly in 2004 plus it has developed policy making and executive bodies, that bring deputy ministers together from national governments. Working groups under the Secretariat are focused primarily on tackling trade, economics, transport, culture, science, education, engineering, energy and combating organised crime and terrorism.125

GUAM's effectiveness is difficult to gauge. One view is that member states may “decide on ‘minilateral’ (small groups with exclusive membership) rather than ‘plurilateral’ (wide membership organisations) relations in order to move forward more easily”.126 Whilst the organisation has, until recently, traditionally found part of its strength through Ukraine providing a type of counterweight to Russia’s regional dominance, a short review of its recent output is illustrative of the coherence of this regional sub-grouping. It must also be acknowledged though that, due to its relative size, the power of its voice is diminished in the international arena to the extent that whilst many eastern European countries and the US have supported it, most EU and NATO members have largely overlooked it.127 In April 2014, GUAM issued a statement backing the UN General Assembly’s Resolution on the territorial integrity of Ukraine; completed working groups on information technology, culture and tourism and transport; hosted a meeting for member’s Foreign Minister’s and jointly analysed the Council of Europe’s most recent report on the conflict in Georgia.128 It is noticeably more contemporary and able to comment on the current regional issues, particularly Ukraine, when compared to BSEC’s apparent reticence to register a view. In the final analysis, strained relations, limited output, the Ukrainian crisis, Moscow’s assertive policies, the West’s indifferent approach and a failure to generate energy cooperation in the face of regional competition,129 all have a debilitating effect on GUAM. One think tank argues that the consequences see the organisation reduced to formal meetings and geopolitically inessential fields of cooperation but it
predicts will continue in its existence.\textsuperscript{130} Avoiding duplication, looking to innovate and leading on a progressive stance on the critical regional issues may be a trump card that continues to stimulate BSEC and wider actor’s activities. GUAM should not be wholly discounted or discredited.

An alternative form of cooperation is offered from outside the region by the EU. It is a significant player, has many attractions and is the main trading partner for all Black Sea countries, effectively becoming “the centre of gravity for the region’s economic activities”\textsuperscript{131}. The EU’s policies are threefold. First the ENP, which was developed in 2002 when the EU was attempting to create a security buffer zone around its borders,\textsuperscript{132} especially in the lead up to sweeping enlargement, including the ex-Soviet Baltic states in 2004. It developed specifically over concerns about the internal situations in Moldova, Belarus and Ukraine that were due to become direct EU neighbours,\textsuperscript{133} but also expanded as a neighbourhood policy to address wider EU regional relations, particularly in North Africa. Second and, in order to provide more specific focus on the EU’s eastern borders, BSS was initiated in 2007. This represents a more forward leaning regional cooperation strategy and effectively acknowledged the Black Sea as a region with a clear political profile.\textsuperscript{134} Its sectorial framework attempts to forge cooperation in democracy and human rights, security, frozen conflict resolution, energy, transport, the environment and a maritime policy including fishery protection, scientific research, trade and social affairs. It is effectively dealing with many of the BSEC members and experience, since its inception, has shown that the greatest benefits have come from partnerships in the transport, energy and environment sectors.\textsuperscript{135} As examined above, these are the fields where BSEC performs best in its own right and the EU synergy plan appears to have added value in specific areas. The third policy is the EaP. It aims to complement the other initiatives and targets members\textsuperscript{136} to provide them with an EU membership perspective, although it is not specifically about accession to the EU nor does it necessarily preclude it.\textsuperscript{137} The European Commission’s president recently stated that any integration must be on a step-by-step basis although it remains a strategic objective. As such, the EaP is “about extending and anchoring stability, rule-of-law, investment opportunities and growth beyond the European Union borders”.\textsuperscript{138}

None of the EU’s initiatives for cooperation are without challenges. Relations with Russia dominate the debate and are being further tested more recently. Critics cite the neighbourhood policy as flawed because it tries to offer a similar deal to a heterogeneous target group and its awkward strategic relationship with Russia.\textsuperscript{139} For instance, the EU has previously supported Russia’s bid to join the World Trade Organisation but now objects to that due to perceived trade restrictive measures, particularly in the automobile industry.\textsuperscript{140} The EaP is criticised as nothing more than an attempt to speed progress on visa facilitations, BSS is potentially its competitor and lacks real synergy or interferes with BSEC processes, and the ENP has not realised any of its formative objectives in a satisfactory manner.\textsuperscript{141} Whichever is more accurate, the challenges remain and the EU’s regional relationships will most likely continue to produce mixed results.

This analysis has not addressed the influence of NATO and the US in any detail but they can have as much of a spoiling effect to counter some of the positive aspects of engagement or expansion. Three of the Black Sea littoral states are NATO members and Georgia remains an enthusiastic and committed aspirant nation with a robust Membership Action Plan, that is due to be reviewed at the next NATO summit this year. Recent events have necessarily forced NATO to reassure its East European allies through increased military air and sea patrols and, whilst they are defensive in their nature,\textsuperscript{142} the chances of misinterpretation also increase as the supra-regional security dilemma plays out. The significant additional factor that has not been addressed is Turkey. A long time NATO member, it is excluded from the EaP despite its long-term candidacy for EU membership and substantial contribution to regional cooperation. Where it stands today is debatable and depends primarily on resolving the division of Cyprus, assessed as requiring a miraculous turn of events in 2012\textsuperscript{143} and overcoming much of the tacit opposition in Europe. Just as the realities and perception of Turkey’s non-European character\textsuperscript{144} obstruct the path to cooperation or potential accession, the same can be said for wider Black Sea cooperation.

Whilst there are numerous schemes promoting integration, the reality is that many are flawed, not fully resourced or supported and lacking in a coherent narrative due to the conflicting requirements of the participants. The fact remains that “for all of the energetic summity that has defined BSEC, it is the
policies of NATO and the EU that are today driving the forces behind the international politics of the Black Sea zone." At the lower levels, best classified under Manoli’s micro-regionalism header, it works and there are tangible results of progress, integration, success and true synergy. As politics, national strategic interests and external actors are introduced it becomes more complicated, strained and demanding. Efforts must continue to bridge the gaps and find the interfaces that can work at the highest levels. BSEC should be the avenue for this but it needs more members to commit to its cause, and not simply use it as a springboard to gain membership to the premium clubs, such as the EU and NATO. Aybak’s analogy of BSEC acting as the “intellectual lifeboat in which to cross the choppy political waters the Black Sea, giving a sense of direction to diplomats, business circles and politicians in the twenty-first century” provides a suitable link to examine maritime security as the final subject for this paper.

**Black Sea maritime security**

“Securing the maritime environment will require the development of comprehensive approaches that involve nationally-focused efforts, allied with regional and global initiatives and cooperation.”

This chapter examines Black Sea maritime security to build on the cooperation theme and explore the practical challenges in the region. After defining the concept, it will outline security threats and assess their potential impact before examining BlackSeaFor and OBSh, to evaluate their capabilities as the principal regional solutions to the problems. The maritime environment is important and provides a natural link to the main areas where BSEC cooperation has been most successful; energy, transportation, tourism, trade and, to a certain extent, protecting the environment. Getting it right, as Sloggett alludes to above, is at first a national issue, particularly in a state’s territorial waters but also a regional one as the sea links the countries. It becomes an even wider issue when addressing environmental concerns as the Black Sea represents a potential marine ecological catastrophe, and is a complex body of water under threat from pollution and a lack of regulation. The sea’s physical connection with the Mediterranean through the Bosphorus, and the fact that the Danube, Dniester, Dnieper and Don rivers all flow into the basin, potentially carrying pollution from central Europe and beyond, suggest that regional initiatives are required as a minimum.

The sea has four historic attributes that make it unique; as a resource and as a medium for transportation, information and the spread of ideas and for dominion. As such, it needs to be protected, must be secure, may be exploited and will be contested. Maritime security has recently been defined in the UK context as the “advancement and protection of … national interests, at home and abroad, through the active management of risks and opportunities in and from the maritime domain, in order to strengthen and extend … prosperity, security and resilience”. The output covers a wide range of operations utilising the full spectrum of maritime forces and their attributes against a mixture of illegal activities as well as protecting the environment. Put simply, it is the “creation and maintenance of security at sea to facilitate prosperity”, but it is not simple to achieve. The International Seapower Symposium recently reported there is value in cooperation and operational initiatives, particularly to develop training opportunities, but there remains information sharing and protection issues and a surmise that a one size fits all approach will not work in the Black Sea. The maritime capabilities and national aspirations of the six countries that border the sea, and can influence it most, varies extremely. This includes Russia’s intent to revitalise her Black Sea Fleet, the balance provided by Turkish southern regional maritime dominance and control of the strategic chokepoint access, and Georgia’s embryonic coast guard fleet under the control of their Ministry of Internal Affairs. It is a diverse combination and any cooperation must be assessed with that caveat. Also, there is much more to maritime security than just naval forces and state business. Whilst outside the scope of this paper, there are considerable implications for the commercial maritime sector and dealing with their vulnerabilities. Maritime security also encompasses port security measures, the cargo supply chain and practical on board shipping protection; a perspective that the primary chokepoints are in fact the sea ports and the ships themselves.

A review of Black Sea maritime history can lead to a conclusion that conflict dominates affairs. Piracy was prevalent during the Ottoman reign and the strategic shifts of the Crimean War, to some extent, hinged on the sea as a neutral international space to curb Russian imperial aspirations. By the end of WW1 the
new Black Sea actors “sought to appropriate the sea’s wealth for their own political, economic and strategic goals” as part of their state-building process. Between wars, the Montreux Convention of 1936 established a legal regime over the Turkish Straits limiting the entry and length of stay of non-littoral state warships and allowing free passage of commercial ships, under the control of Turkey. More recently, in 2000, a Ukrainian coast guard vessel opened fire on and sank a Turkish trawler over fishing rights and territorial water disputes. In 2006, as part of planning for Exercise Sea Breeze, US naval diplomacy shortcomings dented Russian and Ukrainian relations due to the ‘landing’ of troops and equipment ashore to prepare for the exercise without full approval in Crimea, promoting disagreement. Russia blockaded Georgian ports, conducted amphibious operations in Abkhazia and sank Georgian military vessels in 2008 and has recently used similar blockade tactics to constrain the remnants of the Ukrainian navy in Crimea, by sinking an old cruiser at an operational port entrance. More widely, Romania and Ukraine had to turn to the International Court of Justice to resolve disputed claims over their maritime boundaries and the opportunity to exploit gas and crude oil reserves in 2009. Boundaries and national ownership remains contested in a number of areas, for instance, Georgia and Abkhazia’s sea border jurisdiction and border disputes in the Sea of Azov between Ukraine and Russia. Before the crisis, Ukraine’s president had alluded to serious differences in determining the borders and ownership of the Kerch Strait which provides access to the Sea of Azov. The natural resource potential is high although underexplored and will likely cause further tension in the future. In April 2010, Ukraine signed a deal with Russia, the Kharkiv Pact, to extend the Black Sea Fleet’s use of Sevastopol naval base until 2042 in exchange for cheaper gas supply. A precursor to the crisis today, it caused parliamentary unrest and public protest and set the conditions for Russia’s continued dominance in the region. In this context, the maritime environment is characterised by disagreement, dispute and fractured relationships, which leads to a tense situation on and around the sea dominated by ungoverned space and no comprehensive approach to maximise collective prospects.

The situation is potentially worsened considering the array of collective threats that challenge Black Sea maritime security. Generically they include terrorism, piracy, people smuggling, drug trafficking, illegal immigration, criminal activity and threats to the environment from pollution and overfishing. The strategic importance of the Turkish Straits alone means that they must be addressed. The strait is recognised as one of the world’s busiest, with over 50,000 cargo ship transits per year, including over 5,500 oil tankers, and serves as the only means by which some of the Caspian Sea oil exports can be distributed in western and southern Europe. As Black Sea reserves are exploited, and the challenges of the deep water extraction are overcome by international companies who are investing in production sharing agreements, the importance is only likely to increase. Energy independence from Russia for Bulgaria, Romania and Turkey may be achievable in the mid-term, out to 2025, if the exploration trends can be realised. Similarly, while much of the regional energy is transported by pipeline, such as through the BTC or the Blue Stream pipeline that carries gas under the Black Sea from Russia to Turkey, securing the critical nodes from threats is effectively a maritime problem.

Specific threats have emerged outside of the energy sector. In 2000, a commercial cargo vessel arriving in Odessa from Columbia was found to be transporting marijuana in a ‘drugs for guns’ deal between elements of the Russian mafia and Columbian drug cartels. Nuclear materials trafficking is acknowledged as a concern, with reports suggesting that uranium and other radioactive material shipments have been moved using maritime routes in the past, leading to a conclusion that maritime interception and detection capabilities must be improved in the Black Sea. The combined challenges posed by potential terrorist attacks in the region from Caucasus separatist groups, the risks posed by exploitation of illegal activities from organised criminal gangs and arms smuggling to supply the frozen conflict areas and wider conflict zones such as Chechnya, Afghanistan and Iraq, suggest that the pressures exist to demand a solution. Other commentators have identified that some of the deeply entrenched criminal networks, taking advantage of a security vacuum and increasing maritime trade flows, necessitate more constabulary roles for forces at sea rather than just punishing offenders if they are caught. This requires the elimination of the root causes through security sector reform, a reduction in corruption, regional cooperation and economic development to reduce the opportunities for criminal network to flourish ashore and at sea.
On balance, the Black Sea area is relatively safe for mariners and for the protection of critical trade flows on a day to day basis, particularly when compared to global maritime trouble spots. The Malacca Straits are still afflicted by piracy; the Indian Ocean similarly through piracy, smuggling and terrorist threats; and the Mediterranean due to human trafficking from North Africa to Europe. Part of the problem is that, due to a lack of monitoring or the presence of maritime forces, many of the potential threats to security are unlikely to be recognised or occur under the radar of the forces that might have the capacity to deal with them. Civilian port security and practical shipping measures aside, best practice for maritime security requires a coordinated approach to the ends, ways and means of delivery a strategy. Having a strategy comes first (the ‘ends’), effective coordination and control (the ‘ways’) is necessary and a capability to act (the ‘means’\textsuperscript{173}) provides credibility and a deterrence effect.

There are a number of initiatives in the Black Sea to confront the maritime security challenges. These include the Black Sea Littoral States Border/Coast Guard Cooperation Forum (BSCF) which is designed to enhance regional peace, stability and cooperation on maritime security. This small but forward leaning organisation potentially represents a significant step forward at the tactical level of dealing with the threats, particularly through annual exercises for coast guard forces that bring units together for practical training. Also, through the establishment of the Black Sea Border Coordination and Information Centre\textsuperscript{174} in Bulgaria, that aims to coordinate information flow from a variety of sources to provide an integrated real-time maritime picture, taking feeds from a variety of commercial and military sources across the region. The Danube Black Sea Task Force (DABLAS) is a cooperation programme that aims to promote the rehabilitation of ecosystems and pollution reduction in both the Danube River and the wider Black Sea region, including its other river tributaries.\textsuperscript{175} It brings together a wide variety of actors including the littoral countries, interested EU member states and international financing institutions to successfully develop programmes, provide backing and deliver them to achieve its aims. Both the BSCF and DABLAS are examples of effective cooperation within the Black Sea region to deliver smaller programmes, the ways and means, to meet the ends in specific areas of maritime security.

BlackSeaFor, another Turkish led initiative, was established in 2001 to build interdependence for mutual benefits and security at sea. It is styled as a multinational naval on-call task force which aims to revitalise regional cooperation activities, improve relationships between Black Sea countries and increase peace and stability in the region.\textsuperscript{176} Its operational focus and goals are principally centred on “search and rescue operations for humanitarian needs, clearing sea mines, joint action for protecting the Black Sea environment and organising goodwill visits amongst Black Sea countries”.\textsuperscript{177} As a maritime security operation, however, it lacks the required permanent presence to monitor and counter the threats already identified. Commentators have highlighted its limited effectiveness,\textsuperscript{178} but also acknowledge that it does offer a valuable, almost unique, forum for Russian and NATO forces to operate together and act as a confidence-building measure for the Black Sea coastal states.\textsuperscript{179} With a bi-annual activation programme it has, until recently, enabled a basic level of training, the development of interoperability and completed 21 scheduled activations since its inception. Its on-call aspiration, the ability to respond collectively to a threat or incident, has not been tested and any response required at present effectively remains a national responsibility. The task force’s credibility was also damaged during the 2008 war when BlackSeaFor activity continued as planned around Sevastopol, as Russian ships sailed from the base to eventually conduct attacks against Georgia.\textsuperscript{180} At the time, and subsequently, there has never been an official BlackSeaFor comment, suggesting that just as for BSEC, the organisation is limited in its ability to effectively coordinate its members or deal with the more complex areas of regional security. More ominously, the planned April 2014 BlackSeaFor activation has not occurred, again without official word on the reason, although the current events will be a limiting factor. Also Ukraine is soon expected to propose that Russia is excluded from BlackSeaFor,\textsuperscript{181} which has significant implications for the future of the group but is yet to be resolved. In 2013, despite some poor weather, each of the six nations sent a ship to the group for the Spring and Autumn activations and made some progress on cooperation and developing a common computer and communications network for information exchange during exercises.\textsuperscript{182} The overall contribution of BlackSeaFor to regional maritime security has been constrained by the frequency of its activations, the low collective capabilities of the assigned units and its inability to respond as an on-call force to an emerging threat.

221
OBSh to some extent fills the gap. Initiated in 2004 by Turkey as part of a response to the threat of global terrorism, it originally focused on the routine patrols of Turkish naval and coast guard units in their home waters to deter the threat of illegal maritime activities, terrorism and asymmetric threats.\textsuperscript{183} The Turkish authorities extended invitations to the Black Sea nations to join and Russia, Ukraine and Romania have all participated in the initiative to some degree. Due to Turkey’s NATO membership it is closely affiliated with NATO’s Operation Active Endeavour (OAE) in the Mediterranean and, through technology including the Automated Identification System to provide maritime shipping situational awareness and a feed from NATO’s Maritime Safety and Security Information System,\textsuperscript{184} it brings the synergies of OAE’s long standing multinational cooperation and expertise to the Black Sea.\textsuperscript{185} Turkey has established a permanent OBSh HQ and while its progress has been slow, limited again by the naval capabilities of the smaller nations, Russia’s caution over NATO’s growing footprint in the region\textsuperscript{186} and a lack of national enthusiasm to contribute towards safeguarding the maritime environment, it has evolved effectively.

As the two dominant maritime powers in the Black Sea, Turkey and Russia are well placed to build on the OBSh initiative, if some of the critical challenges can be overcome. The current geopolitical environment, regional tension and lack of cooperation are unlikely to support much further advancement in the short term. National capabilities are also a limiting factor. Romania and Bulgaria are slowly developing their naval and coast guard forces and building NATO interoperability, whilst Georgia operates a small coast guard force that is just able to protect its territorial waters,\textsuperscript{187} even though they are disputed. Ukraine’s commitment and capability will depend on the outcome and eventual resolution of their internal crisis. The Black Sea maritime security challenges present a microcosm of the wider regional troubles, both in terms of their impact and the complications of bringing dissimilar nations together to deal with them. Actors from outside the region, such as NATO, the EU and the US, all have vested interests from a global security perspective through to reliance on energy supplies and their influence is potentially destabilising. The answer, at least in the maritime domain, is to build a networked capability, derived from the existing structures, particularly OBSh and BSCF, to identify and confront the threats to safeguard the sea and its users. The corollary is that in the areas that matter to the collective region; trade, tourism, energy, transportation and the environment; the benefits of working together can be realised without overbearing demands. It does require resourcing, investment and better cooperation but has the potential to be a theme where differences are put aside and, at the right levels, synergy is achieved. As noted, one size will not fit all and Black Sea “security cannot be imported, it must be provided by the locals”.\textsuperscript{188}

**Conclusion**

“(The Black Sea) has the potential to become a future pluralistic security community and a region that truly ‘exists’.”\textsuperscript{189}

The Black Sea region is important and recognised as a vital hub at the crossroads of history, culture, politics and energy resources. Its regional definition is contested and, whilst this is really a question of perspective and which lens an observer chooses to view the area, it is conceivable the reason why the Black Sea countries have not yet exploited how they should cooperate. Russia’s part in the complex game of trying to resolve how the region should work together and how to address the frozen conflicts and the wider security issues, including the ongoing Ukraine crisis, is critical. However, it is in the national interests of the Black Sea littoral states and the wider communities to maximise the potential of natural resources, increase trade opportunities internally and externally, and reduce the destabilising effects of conflict, crime and terrorism.

Considering both recent history and current events, regional conflict is the predominant geopolitical feature of the region. Relatively young and arguably weak states are fighting to find their place or seek protection either from the regional hegemons (Russia and, to some extent, Turkey) or within extra-regional organisations (the EU and NATO) and in the context of trying to forge regional structure that provide mutual benefits (principally BSEC). These complicated inter-relationships demand that political skills at the state level, nation building tactics and foreign and domestic policies must be in unison,
working effectively from local to international levels. Conflict, disputes and threats stifle regional options and erode interdependence. Resolving the conflicts must be given more priority, although it is acknowledged that the solutions are most likely years, if not decades, away. This will continue to afflict the region and constrain its ability to move forward collectively.

Black Sea cooperation has been effective in some areas as the paper has highlighted. This provides hope for the future and demonstrates particularly how the ‘micro-regional’ approach may well be the key to unlocking the regional potential. It is possible to summarise that there are too many cooperative initiatives and this again supresses demonstrable progress, certainly at the national state-on-state level. Lower down, and into specific sectors, outside of the multifaceted and intricate energy market demands, cooperation is thriving and countries are beginning to realise how working together to attract tourism, improve transport links and develop trading opportunities can prove mutually beneficial. Whilst BSEC will almost certainly not be the vehicle to resolve the regional conflicts, especially from Russia’s perspective, it should continue to grow in stature and output to ensure its near neighbourhood functions efficiently and with collective purpose while remaining cognisant of the challenges it faces.

At sea many of the regional issues amalgamate. Competition is likely to increase over energy resources as exploration opportunities increase for the large international oil companies and other commentators noting how maritime borders could become a source of conflict. The significant benefits to Russia, now claiming an additional 36,000 square miles of Black Sea ‘territory’ when the Economic Exclusion Zone is applied out to 200 miles from Crimea, will threaten any energy deals and possibly strengthen Russia’s dominance. More widely, just as the physical access to the sea is restricted geographically and then politically via the Montreux Convention impinging on Black Sea maritime activities, there are similar conceptual obstacles to the external influences, such as the EU. Most cooperative initiatives at sea have only generated limited results and the sea, just as a reflection of rest of the region, serves both to divide and also inextricably link the countries that surround it. Resourcing maritime security solutions to the numerous challenges, threats and risks and, in particular, safeguarding the sea itself may be the key to unlock the difficulties of cooperation for the littoral states. If the benefits of such action could flow upstream then the wider region will similarly profit.

Russia’s actions in Crimea and against Ukraine have challenged many of the international expectations and norms of state behaviour, to add a more contemporary context to the deep seated historical challenges of the region. Over time it is likely that the current tensions, bloodshed and disagreement will dissipate and some form of regional normality will be re-established. How much these current events impact on the region is yet to be determined but, in order to survive, the Black Sea must become an area of future cooperation and ultimately solve its conflicts. This is an idealistic view, but one shared by many commentators, as Weaver notes above. The sea is as good a starting point as any and, if the maritime community collaborates, it allows them to “develop a coherent, collaborative framework that spans national boundaries, and become agile and responsive…to deter and disrupt…those that threaten their livelihoods”. This is a practical blueprint for the countries themselves.

Endnotes


King, Black Sea, 192.


Carol Weaver, The Politics of the Black Sea Region: EU Neighbourhood, Conflict Zone or Future Security Community? (Farnham: Ashgate, 2013), 140.


ibid., 430-431

ibid.

Ascherson, Black Sea Region, 42.

ibid.

ibid.

ibid.

Ascherson, Black Sea, 50.


ibid., 29-30.

ibid., 30.

ibid., 111.


King, Black Sea, 191.

King, Is the Black Sea a Region?, 19.

For details see King, Black Sea, 224-227.

ibid., 227.

Albania, Armenia, Azerbaijan, Bulgaria, Georgia, Greece, Moldova, Romania, Russia, Serbia, Turkey and Ukraine.

Panagiota Manoli. The Dynamics of Black Sea Subregionalism (Farnham: Ashgate, 2012), 1-3.

Bulgaria, Georgia, Romania, Russia, Turkey and Ukraine.


ibid., 6 and 16.


Manoli, Black Sea Subregionalism, 135-138.


ibid., 16.


Weaver, Black Sea Region, 89.


Weaver, Black Sea Region, 89.


ibid., 272-273.

59 Weaver, Black Sea Region, 97.
60 ibid., 96.
61 Medar and Savu, Frozen Conflicts in the Black Sea Area, 44.
64 Paul Mason, “If Transnistria is the next flashpoint between Putin and the west, how should Europe react?,” The Guardian, April 1 (2014), http://www.theguardian.com (accessed April 2, 2014).
66 ibid., 275.
67 Ascherson, Black Sea, 247.
68 ibid.
69 ibid., 249.
70 Weaver, Black Sea Region, 80.
71 Jones, Georgia: A Political History, 236.
74 The UN, EU, OSCE, US, Russia, Georgia and representatives from South Ossetia (and Abkhazia) attend. For details see Weaver, Black Sea Region, 85-86.
75 For details see Gahrton, Georgia: Pawn in the New Great Game, 177-191 and Jones, Georgia: A Political History, 240-245.
77 ibid.
78 De Waal, Caucasus, 215.
79 Weaver, Black Sea Region, 85.
80 Jones, Georgia: A Political History, 244.
84 ibid., 171.
85 ibid., 102.
88 Reid, Borderland, 187. Potemkin allegedly built a false town frontage in the Crimea to impress Catherine the Great during her visit in 1787.
89 ibid., 233.
91 ibid.
92 King, Black Sea, 242.
95 Manoli, Black Sea Subregionalism, 77.
99 Manoli, Reinvigorating Black Sea Cooperation, 9.
100 ibid., 13.
101 Manoli, Black Sea Subregionalism, 43.
102 ibid., 43-44.
103 ibid., 57.
110 Weaver, Black Sea Region, 127.
111 ibid., 128.
112 BSEC Charter, Article 20.
113 Weaver, Black Sea Region, 127.
114 Hajizada, Complex Regionalization in the Black Sea Area, 122.
116 Manoli, Black Sea Subregionalism, 97.
117 Weaver, Black Sea Region, 128.
118 ibid., 129.
120 Manoli, Black Sea Subregionalism, 98.
122 Ibid.
124 Manoli, Black Sea Subregionalism, 51.
126 Hajizada, Complex Regionalization in the Black Sea Area, 122.
127 Manoli, Black Sea Subregionalism, 54.
129 Particularly from Turkey as the central hub of regional energy corridors and see Manoli, Black Sea Subregionalism, 54.
131 Manoli, Reinvigorating Black Sea Cooperation, 16.
132 Weaver, Black Sea Region, 45.
134 BSS engages with Armenia, Azerbaijan, Georgia, Moldova, Russia, Turkey and Ukraine.
136 EaP includes Armenia, Azerbaijan, Belarus, Georgia, Moldova and Ukraine.
137 Weaver, Black Sea Region, 50.
141 Cichocki, European Neighbourhood, 26.
143 Park, Modern Turkey, 63.
144 Ibid.
145 King, Black Sea, 244.
149 For details see Ascherson, Black Sea, 257-272.
155 For detail on seaport and ship security see Michael McNicholas, Maritime Security An Introduction (Burlington: Butterworth-Heinemann, 2008), chapters 8-10.
156 King, Black Sea, 125 and 141.
157 Ibid., 190.

159 King, Black Sea, 216.


169 McNicholas, Maritime Security, 259.


171 Sanders, Maritime security in the Black Sea, 104-107.


174 Manoli, Black Sea Subregionalism, 48.


177 ibid.


179 Manoli, Black Sea Subregionalism, 47.

180 ibid.


183 Sanders, Maritime security in the Black Sea, 108.

184 An internet enabled, multi-source maritime intelligence hub.


189 Weaver, Black Sea Region, 140.


How do we end piracy on the high seas?

SQN LDR M D LOCK RAF

Introduction

“International shipping transports about 90 per cent of global trade to peoples and communities all over the world. Shipping is the most efficient and cost-effective method of international transportation for most goods; it provides a dependable, low-cost means of transporting goods globally, facilitating commerce and helping to create prosperity among nations and peoples.”¹

Maritime piracy can damage the sea trade that the world relies on and can cost the global economy up to £7.6 billion a year.² Although this significant cost is calculated using first-order expenditures, such as the cost of ransom payments, on-board deterrence measures and naval costs, it does not reveal the second order impacts, such as damage to foreign investments or tourism or commodity price inflation³ that can be far more reaching. Consequently, it is very difficult to gain a full understanding of the impact of piracy with so many variables. The high number of variables also means that the effectiveness of mitigations and solutions is also difficult to determine.

The issue of piracy is one that is dealt with under international law, but the extent to which it is dealt with will, inevitably, be influenced by resources and political will. Piracy, by definition, is a crime that is limited solely to the high seas, which is defined as “all parts of the sea that are not included in the exclusive economic zone, in the territorial sea or in the internal waters of a State, or in the archipelagic waters of an archipelagic State.”⁴ Different interpretations of the term piracy will affect the statistical analysis of the problem, which will again cause difficulties when trying to determine the extent to which piracy is currently an issue.

In order to examine how the international community could end maritime piracy, this paper will consider a number of different perspectives. Although a great deal of effort is currently focussed on combatting piracy operations at sea, with a better understanding of the foundations, support and funding of the crimes, it can be possible to more fully appreciate the ways to tackle it. An improper understanding of what causes piracy may lead to inefficient and incorrectly targeted interventions. There is also the possibility that a lack of understanding could lead to the legitimizing of piracy as a source of employment in the minds of the local population.⁵ It is also important to have a comprehension of the scale of the problem to determine the level and type of resources that should be used to combat piracy, and to determine if it is really a problem that should be treated, or if it is one that can just be tolerated.

The United Nations Convention on the Law Of the Sea (UNCLOS) provides the framework for the control of piracy under international law. UNCLOS details that all states shall have freedom of navigation, but this is a freedom that is restricted when pirates carry out their crimes. Piracy itself is covered under Article 101 of UNCLOS and consists of the following acts:

“(a) any illegal acts of violence or detention, or any act of depredation, committed for private ends by the crew or the passengers of a private ship or a private aircraft, and directed:

(i) on the high seas, against another ship or aircraft, or against persons or property on board such ship or aircraft;

(ii) against a ship, aircraft, persons or property in a place outside the jurisdiction of any State;
(b) any act of voluntary participation in the operation of a ship or of an aircraft with knowledge of facts making it a pirate ship or aircraft;

(c) any act of inciting or of intentionally facilitating an act described in subparagraph (a) or (b).”

Whereas piracy takes place on the high seas, similar crimes that take place within a state’s internal waters, archipelagic waters and/or territorial sea are widely recognised and defined as acts of armed robbery against ships. Throughout this paper, these acts of armed robbery will be included when referring to piracy, as the crimes are very similar and, in general, affect seafarers in the same way: for example, through the risk of injury, death, kidnap or loss of possessions or cargo and in terms of insurance and risk mitigation and the preventative measures which can be employed.

The nature of piracy differs from region to region due to the varying socio-economic circumstances of the perpetrators, the level and condition of governance within the communities and states and the preventative measures put in place by the shipping industry, local law enforcement and the international community. The international community’s efforts to tackle the regional piracy problems, particularly around the Horn of Africa, are both vast and costly. Although these efforts have produced quantifiable results, they do also show that ending piracy on the high seas will never have a 'one size fits all solution'. Piracy is like any other organised crime in that it is hard to eradicate due to the will of the criminals involved. Additionally, the criminals themselves will evolve in nature depending on the rewards and opportunities they encounter, and the resistance they face from those who wish it to stop. Piracy, in particular around Somalia, has evolved into a transnational operation which started as small, locally executed and resourced ventures into wide networks with funding, equipment and people coming from abroad, and proceeds being reinvested in further operations or in people trafficking and other organised crime.

In addition to the perpetrators who carry out the piracy acts, and the state forces who try to prevent piracy, there are also a myriad of businesses, enterprises, corporations and industries that have been established to tackle it from both a security and a financial perspective. Therefore, eradicating piracy completely would put the multi-billion dollar anti-piracy industry out of business so it is argued that doing so would not be in their interest. The preventative measures, security sector and insurance industries will be considered later on in this paper, along with an assessment of the financial benefit that they attain as a result of the crime. Measures which shipping companies and seafarers can consider to mitigate the risk of piracy will also be outlined and the paper will show that, although some of these can offer additional protection and deterrent effect, they can also bring with them inherent risks and costs.

The aim of this paper is to examine whether the international community can put an end to piracy on the high seas. Piracy is a lucrative business but can only be carried out when the conditions are right. Therefore, to begin with, this paper will consider the means, motives and opportunity that are required to set the conditions to enable piracy. It will then explore the nature of modern-day sea piracy, identify trends across the hotspots of the Straits of Malacca, the Gulf of Guinea and the Gulf of Aden, make an assessment of the impact of piracy on the global economy, in terms of risks and opportunities, and then assess the role of the international community in combating future piracy at sea and on land. From Diagram 1 it can be seen that other areas, such as the Indian Ocean, the Mediterranean Sea and around the coastline of South America, are also affected by piracy, however, this paper will concentrate its analysis around Africa and the Straits of Malacca, as this is where the vast majority of piracy currently takes place. This paper will conclude that mitigation and further reduction of the occurrences of piracy is possible, but it will never be eradicated as currently the rewards for those involved far outweigh the risks. There can be an economic reliance on the pirates within communities and also on the legitimate piracy business of insurance and prevention so it will not be in everyone’s interest for it to end. Furthermore, as there is currently a status quo where the maritime industry as a whole is not losing enough money for it to become a significant factor, and the risks for crews and pirates are low enough for them to carry on, there is not the real incentive for any party to try to stop the activity altogether.
Motives

Piracy can be a form of organised crime and the motives for its origination can be attributed to the economic conditions which the pirate communities find themselves in. Two of the three piracy hotspots, the Gulf of Aden and the Gulf of Guinea fall within areas which are fragile or failed states. These areas have high levels of lawlessness, and have weak legal and institutional structures that cannot fully provide security and infrastructure to their populations. This lack of authority is widely acknowledged to be one of the root causes of the development of piracy from traditional fishing communities. Where the communities used to rely on fishing as a means of both providing food and a source of income, the increase in the globalised nature of fishing meant that large fishing vessels and fleets from as far afield as France, Spain, Japan and Korea exploited and devastated the unprotected local fishing grounds. International and Somali observers have stated that up to 700 foreign industrial fishing vessels at a time were involved in illegal and unregulated fishing in the seas around Somalia, frequently engaging in intentional collisions with local fishermen. The foreign fleets destroyed fishing equipment and severely depleted fish stocks that the Somali fishermen relied on. It is estimated that this illegal and unregulated activity over the past twenty years could account for over 50% of the overall catch around the unmonitored waters off Somalia, and that the fish taken illegally by foreign fleets could have been worth as much as $300 million. These estimates indicate some of the damage that foreign fleets have caused to the local economy due to a lack of effective policing.

Without an effective government to either police the fishing grounds or challenge the perpetrators in front of the international community, the practice of illegal fishing went unchecked and unchallenged which left the local fishermen with two possible courses of action. The first was to challenge the foreign fleets themselves, by either trying to deter them or by requiring them to pay an unofficial levy for a ‘licence’ to allow them to continue to exploit Somalia’s marine resources. This could partly compensate the fishermen for the loss of their livelihood, but in reality would mainly fund corrupt landlords and militia leaders. The other option was for the individuals to seek alternative employment and sources of income, but in a fragile or failed state this could prove difficult with little chance of any centralised support to help them. With extremely high levels of poverty and unemployment, many of the former fishermen were forced to turn to crime in order to sustain themselves and their families. As the fishing industry declined through lowered stocks or because of the risk of fishing in the dangerous waters, the
fishermen became “an easy target for pirate leaders and financiers willing to develop their activities” and many individuals found themselves turning to this crime.

Once in the business of piracy, it can be difficult for individuals to escape after they become accustomed to the lavish lifestyle that the money brings, and also because of petty fines that gangs can impose on their crews. It is reported that fines can be imposed on gang members by their leaders of up to $10,000 for mistreating a crew, falling asleep on duty or refusing to follow orders. These fines can quickly turn into massive debts meaning pirate foot soldiers are stuck in a vicious circle, where they have to participate in more raids to pay back the money they owe.

In addition to the loss of livelihood caused by over-fishing, it is understood that a significant number of tankers passing through the Gulf of Aden have washed out their holds directly into the sea creating large areas of pollution. This pollution would damage the marine environment making fishing more difficult due to dying fish stocks, and as a result of the hazards of operating in a polluted sea. Again, without a state voice to protest or prosecute the perpetrators the offence can carry on unchallenged. Information about this practice stems from reputable sources and related data and indicates that illegal dumping, including radioactive and medical waste, has been going on for almost twenty years. The allegations, however, remain difficult to prove due to the security situation in Somalia. In an effort to help counter the problem the UN has considered the accusations, and has included an action within UNSCR 2020 for nations to “positively consider investigating allegations of illegal fishing and illegal dumping, including of toxic substances, with a view to prosecuting such offences when committed by persons under their jurisdiction.” How far resolution 2020 will go towards improving the livelihoods of Somali people is yet indeterminable, but so far these issues are being overshadowed by the efforts to tackle the pirates themselves rather than these alleged causes.

Large ransom payments provide strong motivation for involvement in piracy. In April 2005 a Hong Kong based company is understood to have paid $315,000 for the release of the vessel Feisty Gas, a LPG tanker, and in September 2008 $2 million was paid for the release of the Malaysian tanker Bunga Melati 5. In the region of the Horn of Africa, it is estimated that between $339 million and $413 million was paid in ransoms between April 2005 and December 2012 as a result of piracy. Pirates at the bottom end of the system may receive a payment of between $30,000 and $75,000 (which would amount to about 0.01–0.025 percent of an average ransom payment). Considering that the average income for a Somali national is only around $2 a day, it is easy to see the incentives for the individuals to get involved.

Means

The means to conduct piracy can be considered as the equipment and manpower required to carry out the piracy operation, and the environment to undertake it in free from persecution by authorities. The equipment needed is generally freely available as all that is required is a seaworthy craft, small arms and enough willing personnel to be able to board a vessel. The manpower required will vary depending on the nature of the piracy to be carried out, which may range from stealing possessions of crewmembers, stealing cargo, including ‘bunkering’ of crude oil, holding the crewmembers hostage or taking the vessel itself as a means for ransom. Should the pirates have the aim of taking hostages it is likely that they will need a facility to keep the hostages secure as well as personnel to guard them. Negotiators may be employed to mediate between the pirate gangs and the insurance or shipping companies in order to obtain a ransom payment, and lawyers and translators may be involved on both sides. Piracy is now so well established that all of these facilities and services are readily available and provide the accessible means to carry out the crimes.

Like most other organised crime operations, piracy is built around a hierarchical structure. At the top are the financiers and organisers, at the bottom are elements of the local economy and local private entrepreneurs who support and supply the pirates with goods and services. The vast majority of ransom payments will go to the pirate financiers and this money can be re-invested into other piracy operations or it could go towards other organised crime activities such as prostitution, drugs or money laundering. Due to the increasing amount of money that is changing hands from ransom payments, pirates’ financiers and gangs are developing more complex and robust methods to manage the money. If pirates
are paid in cash, gangs can experience difficulties moving the physical volume of money without being caught, although the smuggling of cash into and out of the porous borders of Somalia is relatively easy. With the rapid increase in mobile technology across the region, electronic financial transfers are becoming more accessible and widely used to manage payments. However, although easier, electronic transactions leave an audit trail which can be exploited by law enforcement officials who have identified Djibouti, Kenya, and the United Arab Emirates as three of the main transit points for financial operations into and out of Somalia, and it is leads like this that can help to cut off the support for pirate gangs at source.

**Opportunity**

The methods by which pirates carry out their crimes are widely documented and by examining these methods, countermeasures and tactics can be developed to oppose them. The opportunities for the pirates and their methods of operation tend to vary regionally and attacks can take as little as fifteen minutes to carry out. In the area of the West Coast of Africa, oil exports through the Gulf of Guinea can be a particularly lucrative target and pirates may bribe corrupt government or industry officials in order for them to reveal when shipments are due to be made or ships loaded. Ships which are at anchor to load oil and to do business are particularly vulnerable which makes them easy targets for criminals. A common thread of opportunity is evident across piracy hotspots with many considering “piracy as a viable economic activity [which gives] economic mobility for young unemployed men.” The nature of piracy around the West coast of Africa tends towards the seizure of cargo, rather than exploitation of hostages, therefore, statistics from the International Maritime Bureau (IMB) reports show that West African pirate attacks involve the use of a greater level of violence as there is no requirement to obtain a ransom and no financial benefit for treating a crew well. The main threat is to vessels doing business with the littoral states and, therefore, there is often little choice for the crews to keep clear of the coast to avoid the threat. Although oil bunkering is a significant problem in the Niger Delta region, and there may be evidence to suggest that Nigerian pirate groups are drawing resources from oil theft networks, this paper will avoid focussing on bunkering. There may well be similarities in the ways to end piracy and ending bunkering at sea, but the bunkering easily blends into the land-based theft of oil and associated corruption that is prevalent in Nigeria which will need a different approach to counter.

In the Horn of Africa area, piracy is a well-known threat, so to mitigate the risk vessels may choose to keep clear of the region, which means pirates have to adjust their methods to create opportunities to attack. Rather than sending out small vessels or skiffs direct from the shore to loiter or carry out the attacks, pirates may steal a larger craft and use it as a ‘mother ship’ where pirate gangs can wait with their smaller craft. Once a suitable vessel has been identified to attack, via listening in to radio chat, using a ship’s Automatic Identification System or radar, the pirates can proceed from the relative security of the mother ship. These attacks tend to be further out from the coast which allows pirates to have access to greater numbers and varieties of shipping. Although the mother ships are less stealthy than the skiffs, once hostages are on board the pirate gangs have a valuable means of bargaining to prevent their capture by navies and enforcement vessels.

In the Straits of Malacca, pirates generally use their attacks as an opportunity for stealing a vessel’s cargo or valuables or taking hostages for ransom. The methods used to identify targets and to attack them can be very similar to those used by Somali pirates, and similarly, as many of the pirates originate from fishing villages they can also retreat back to the shore and disappear into their communities after an attack, often with corrupt authorities turning a blind eye. Most instances of piracy recorded in IMB reports for South East Asia are simple opportunistic attacks and robberies and, as most crews do not physically oppose them, violence is rarely used.

As mentioned previously, it is recognised that well planned attacks can take as little as fifteen minutes, so deterrence must be one of the key weapons in controlling the opportunity to conduct attacks. Pirates who have been interviewed have reported that they have given up on proceeding with an attack because they have been deterred by the presence of watchmen, which in some cases turned out to be dummy silhouettes. Deterrence is crucial, as once a pirate gang boards a vessel authorities may
consider it too risky to intervene, as the result may be that either the gang may turn on the crew or that crew members may get caught in the crossfire during a rescue attempt. Therefore, the numbers of patrols available to deter, the speed of the patrol’s reactions to distress calls and the deterrent measures that a crew can employ will all influence the opportunities for whether a pirate attack will be successful or not.

Understanding modern historical trends

Having considered some aspects of the means, motives and opportunities for piracy, this paper will now consider the extent to which piracy is a globalised problem so that an assessment can be made on the need to end piracy. In order to assess the extent of the piracy problem it is useful to look at reporting and statistics. The UN’s International Maritime Organisation (IMO) and the International Chamber of Commerce’s IMB both produce a breadth of data that can be studied to determine the magnitude of the problem. Although piracy is widely reported as being a serious threat to international trade and shipping, the statistical reality puts this reporting in a different perspective. A report by the IMB on 15 January 2014 said “Piracy at sea has reached its lowest levels in six years, with 264 attacks recorded worldwide in 2013, a 40% drop since Somali piracy peaked in 2011.” This indicates that the measures currently being undertaken by the international community and the maritime industry are making a big impact. However, this drop comes after some difficult times for seafarers. In 2009, Lord Levene stated that “we are experiencing the highest levels of piracy risk for four centuries.” Statements like this may lead a reader to consider that the seas are infested with pirates and the risk of attack is high. However, when the estimated $7.6 billion annual cost of piracy is compared against the losses of cargo on land, which have been estimated to be in the region of $30-50 billion, it puts the problem in some perspective. Although piracy data varies from institution to institution, this paper does not consider it necessary to compare one set of reporting against another or to analyze the reasons for differences. Instead, it contends that an understanding of the definitions used in the reporting of statistics can help to get a better understanding of what is actually being reported.

The UNCLOS definition of piracy is based on acts on the high seas, whereas a large number of acts of seizure of vessels, crew and cargo take place within territorial or other waters and, using the UNCLOS definition to bound the statistical analysis would potentially skew the perception of the problem. For example, within the Straits of Malacca, which varies in width from 200 miles to 11 miles, a large proportion of the navigable area is either designated as territorial waters (including ports and harbours), exclusive economic zone, archipelagic or internal waters. Therefore, acts that are traditionally understood as being piracy that take place in the Straits of Malacca would be excluded statistically by using the UNCLOS definitions. Although it is important to recognise that the statistics may vary depending on the bounding of definitions, they will also be varied by the level of reporting by shipping companies who may wish to underplay the true extent of piracy, as it could result in an increase in their maritime insurance rates or, alternatively, reporting an incident could lead to a vessel being laid up for long periods whilst an investigation is carried out. The IMB claim that as many as half of all pirate attacks go unreported which means an accurate assessment of the true extent of the problem is virtually impossible.

The IMB used to combine piracy and armed robbery at sea for statistical purposes and used its own definition of piracy up until 2009. However, from 2010, the IMB aligned itself with the IMO’s definitions, which will have influenced the data available when considering the extent of the problem over the past decade. Despite these inconsistencies, the data that is available does give us a useful indication of the trends, types and quantities of attacks for which an assessment can be made on the extent to which piracy is a global issue and how it should be addressed. From Table 1 it can be seen that over the past ten years the global trend, based on the IMB’s reported statistics, is for piracy attacks to be on the decrease from a peak around three years ago. The table also shows that being taken hostage is still a significant threat if a vessel is attacked and that being either killed or injured during an attack is rare.
Despite the global economic downturn, reliance on sea routes is on the increase and, therefore, statistically this could provide more opportunities for pirates. World seaborne trade grew by 4% in 2011 whereas the world shipping fleet between 2008 and 2012 grew by 37%. Considering the Gulf of Aden piracy area in isolation, approximately 20,000 ships transit the area that connects the Mediterranean Sea to the Red Sea each year alone, therefore, the statistical probability of an attack can be shown to be quite low in that area. Table 2 compares the number of reported attacks and attempted attacks in selected areas over the past ten years, and shows that the number of vessels affected that were transiting from the Red Sea through the Gulf of Aden was only fifteen in 2013. Even with an inaccuracy estimation of 100% to account for unreported attacks, this still only amounts to a 0.15% chance of an attack or an attempted attack in 2013 in the area of the Gulf of Aden. Using the same assumptions for calculations, this then compares to a 2.36% chance of attack in the same area in 2011 when piracy was at its peak. Therefore, using the data from this table, and using comparable shipping numbers elsewhere, globally the current risk from piracy can be shown as being relatively low, particularly when also considering that these attacks take place over hundreds of thousands of square miles of ocean.

Table 1. Types of violence to crew

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Taken hostage</td>
<td>359</td>
<td>148</td>
<td>440</td>
<td>188</td>
<td>292</td>
<td>889</td>
<td>1050</td>
<td>1174</td>
<td>802</td>
<td>585</td>
<td>304</td>
</tr>
<tr>
<td>Kidnap/Ransom</td>
<td>0</td>
<td>86</td>
<td>13</td>
<td>77</td>
<td>63</td>
<td>42</td>
<td>12</td>
<td>27</td>
<td>10</td>
<td>26</td>
<td>36</td>
</tr>
<tr>
<td>Crew threatened</td>
<td>65</td>
<td>34</td>
<td>14</td>
<td>17</td>
<td>6</td>
<td>9</td>
<td>14</td>
<td>18</td>
<td>27</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Crew assaulted</td>
<td>40</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td>29</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Crew injured</td>
<td>88</td>
<td>59</td>
<td>24</td>
<td>15</td>
<td>35</td>
<td>32</td>
<td>69</td>
<td>37</td>
<td>42</td>
<td>28</td>
<td>21</td>
</tr>
<tr>
<td>Crew killed</td>
<td>21</td>
<td>32</td>
<td>0</td>
<td>15</td>
<td>5</td>
<td>11</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Missing</td>
<td>71</td>
<td>30</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>21</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>644</td>
<td>401</td>
<td>509</td>
<td>317</td>
<td>433</td>
<td>1011</td>
<td>1167</td>
<td>1270</td>
<td>895</td>
<td>662</td>
<td>373</td>
</tr>
</tbody>
</table>

This table only includes data from selected geographical areas and does not account for all reported actual or attempted attacks.

Table 2. Number of reported actual and attempted attacks per geographical area

<table>
<thead>
<tr>
<th>Location of Attacks</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indonesia/Malacca Straits/Singapore Straits/Malaysia</td>
<td>156</td>
<td>149</td>
<td>101</td>
<td>76</td>
<td>62</td>
<td>46</td>
<td>42</td>
<td>53</td>
<td>74</td>
<td>101</td>
<td>125</td>
</tr>
<tr>
<td>Gulf of Aden/Red Sea/Somalia</td>
<td>21</td>
<td>10</td>
<td>45</td>
<td>20</td>
<td>44</td>
<td>121</td>
<td>212</td>
<td>217</td>
<td>236</td>
<td>75</td>
<td>15</td>
</tr>
<tr>
<td>Congo/Chad/Nigeria/Ghana</td>
<td>42</td>
<td>33</td>
<td>19</td>
<td>15</td>
<td>46</td>
<td>42</td>
<td>34</td>
<td>22</td>
<td>16</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>Totals</td>
<td>219</td>
<td>192</td>
<td>165</td>
<td>111</td>
<td>152</td>
<td>209</td>
<td>238</td>
<td>302</td>
<td>326</td>
<td>207</td>
<td>172</td>
</tr>
</tbody>
</table>

This table only includes data from selected geographical areas and does not account for all reported actual or attempted attacks.

Table 2. Number of reported actual and attempted attacks per geographical area

If the number of globally reported attacks is compared to the number of vessels that made trips, it can be considered that, statistically, piracy is not a significant threat to either the ships, their cargoes or the crew members on board. However, the risks must be recognised as they are real and, without this acknowledgement, it is unlikely that crews would be willing to put themselves at risk without more financial reward, which would then lead to increases in commodity prices as the cost of shipping would rise. If the risk of piracy was considered in monetary terms, it has been assessed that less than 2% of the costs that are attributed to dealing with piracy actually ends up in the hands of the pirates and, of this, only a third may be distributed amongst the foot soldiers whereas the rest would go to financiers, logistics, bribes to officials and other operational expenses. The other 98% of the expenditure goes towards increased insurance premiums, Private Maritime Security Companies (PMSCs), anti-piracy
measures on ships and deterrence such as naval patrols. The protection costs borne by shipping companies are ultimately passed on to the consumer of the shipped goods, whereas the deterrence costs are either passed onto the tax payers of the contributing nations or through nations’ contributions to NATO or the EU.\textsuperscript{43} Using this analysis it can be reasoned that there is ultimately a financial benefit for some states involved in counter-piracy operations in terms of legitimate business interests gained from piracy.

**The risk versus rewards for pirate gangs**

Having demonstrated what the risks are to seafarers, in terms of the likelihood of an attack, this paper will now consider the risks and rewards from the perspective of the pirate gangs in order to assess whether current tactics to end piracy are credible. If an attack is successful and a ship is boarded, the benefits for the pirate gangs can be considerable and can far outweigh the judicial risks or the risks of injury during an attack that they may face. Costs can be as little as $300 for a small scale piracy operation where a single family operates from a small craft, to as much as $30,000 for a large syndicate that has two or more vessels in its inventory. Countering this ease of access would need a firm deterrent but this does not appear to be in place. In the Somalia area, the EUNAVFOR report that, since 2009 they have only transferred 154 pirates to competent authorities for prosecution and, of these, 33 were remanded and only 121 have been convicted.\textsuperscript{44} It is unlikely that many of the pirates will fear Western naval patrols in terms of risk as once they are arrested the chances of ending up in jail are currently remote. Due to human rights and custody regulations it is likely that detained pirates will also benefit from a shower, hot food, clean clothes and a medical check-up prior to being moved on or released. The risk of losing weapons or a boat once caught will affect small time operators more than larger ventures as it would remove their means to operate. Larger ventures are more likely to be financed by criminal networks and this loss would be of less impact, as the resource would be available to start up another operation quickly again.

With low risk, the pirates can also face a high reward. Between 2005 and 2011 there was a steady increase in the average ransom payments to Somali pirate gangs, which was estimated to rise from $0.39 million to $5.04 per vessel,\textsuperscript{45} although these figures are very approximate and the actual payments vary considerably depending on the gang, the negotiators (on both sides), the number in a crew and the area in which the hostages and ships have been captured. The 2012 Somalia Human Development Report lists Somalia as having an estimated per capita income of $284, the fourth lowest in the world. Furthermore, the unemployment rate for individuals aged between 14 and 29 is 67%, one of the highest rates in the world.\textsuperscript{46} With ransom demands that can run into millions and few alternative options to make a living, it is easy to understand why the ‘business of piracy’ is so attractive and why so many gangs have been formed in the past.

Although piracy can be a source of income which can boost a local economy, it can also have a destabilising effect. Efforts to develop effective governments and institutions are likely to be undermined by the pirate leaders and financiers in order to protect their positions of power and their investments. Therefore, it is likely that violence, intimidation and corruption will be used to oppose the development of the communities or states as the introduction or development of law and order would threaten the privileged lifestyles of those at the top of the pirates’ food chain.

**The risk versus costs for the shipping industry**

The shipping industry has one of the largest interests in ending piracy on the high seas and faces a number of dilemmas in terms of risk and cost. Few governments openly subscribe to policies of paying ransoms, as such a posture is likely to encourage pirate gangs to exploit the situation, and it is likely this would lead to an increase in pirate activity. In February 2012 Prime Minister, David Cameron, called for the creation of “an international taskforce on ransoms...[with] the ultimate ambition of stopping these payments because in the end they only ensure that crime pays.”\textsuperscript{47} Although this position was fully supported by US Secretary of State, Hillary Clinton, it was not supported by four of the large international shipping companies present at the conference who wrote to David Cameron stating that:
“any action taken by Governments to discourage the payment of ransoms to pirates ignores the plight of those being held hostage and those that risk their freedom and their life daily working on ships in the region. The shipping industry strongly maintains that paying a ransom is the only reliable method for securing the safe release of captured vessels.” 48

Although states rarely support ransoms, the legal position of their payment can remain unclear as was shown in the 2010 case where the English High Court and the Court of Appeal ruled in the case of Masefield v Amlin that, “ransom payments to pirates are not against English public policy and [are] therefore legal.” 49 A government’s position may be at odds with individuals, institutions or shipping companies who are likely to be keen on paying the ransoms as their interests are in the immediate release of the hostages or ships, not the shaping of future policy or management of future risks. “Ransom payments have come to be seen by some ship owners, as a cost of doing business.” 50 Evidence of the extent of ransom payments by non-government actors is difficult to obtain as the payments are normally made covertly, for the same reasons most states avoid ransom payments. Additionally, with a state advising a posture of non-payment, non-state actors would not want to be seen as contravening this advice when a state may also be providing negotiators and diplomatic assistance in order to free hostages and should they find out a ransom has been paid, governmental services could be withdrawn prior to the release of a hostage or vessel. Therefore, although the paying of ransoms is costly in terms of money, they are currently seen by the shipping industry as the only reliable means of ensuring the safe release of hostages and vessels and are seen as a necessity, even if they perpetuate the crimes that put their people in harm’s way. 51

In addition to the ransom costs of piracy, the cost to industry is also borne in a number of other areas which are all eventually passed on to the consumer. Insurance premiums alter depending on the area a vessel will operate in or transit through and the premiums will rise significantly if an area is identified as one where there is a risk of a pirate attack. This rise in premiums leaves shipping companies with a number of options, the majority of which have a direct cost to business: insure the vessel; don’t insure the vessel and take the risk; provide preventative measures to try to reduce the premium; change the speed of the vessel when transiting areas of risk or seek alternative routes.

The International Chamber of Shipping (ICS) has estimated that the additional payments for insurance due to Somali piracy alone is between $460 million and $3.2 billion per year. 52 These risk policies rose 300 fold from 2008 to 2010 from $500 per ship per voyage to $150,000. 53 In addition to the insurance costs, there will also be financial consequences for shipping companies if a vessel if captured as it is likely to be out of service for a long period of time. The international law firm Ince and Co tracks the length of time that vessels are held captive, and has calculated that the average period held between February 2010 and February 2011 was 214 days, a significant increase from the 93 day average the previous year. 54 This further demonstrates the importance of ransom payments from the shipping industries’ perspective, as prompt payment is likely to mean that vessels are back in operation and generating income sooner.

The three main centres of piracy activity: the Gulf of Aden and Horn of Africa; the Gulf of Guinea and around South East Asia also contain the busiest sea trade routes in the world. Therefore, an option of changing routes to avoid the piracy risk can be difficult. As the Gulf of Guinea is one of the main sea export routes for oil from Nigeria there are no alternatives for ships but to use these waters. The risk posed to the ships is less than the financial and human cost of piracy and, therefore, multinational corporations rarely consider halting their operations, as there is no shortage of potential carriers for their goods. The ICS estimated in a 2010 report that the total costs of re-routing ships to avoid the Gulf of Aden is between $2.4 to $3 billion per year. 55 Clearly this amounts to a sizeable addition to the shipping industry’s overheads and will particularly affect low and slow vessels that are most vulnerable to piracy attacks and may choose to re-route. 56 Avoiding the Gulf of Aden by routing around the Cape of Good Hope will inevitably lead to increases in fuel and labour costs and vessels would be able to make less journeys per year because of the time penalties of travelling further. However, choosing this option has the benefit of lower risk, lowered insurance premiums and avoids the Suez Canal charges, which for ships can range from $200,000 to $600,000 each way depending on their size. Therefore, an extensive cost-benefit analysis has to be undertaken by shipping companies to balance the costs against the risks, 58 but this must also be considered against consumer models of ‘just in time’. Many companies now only stock
the goods that they need to keep storage costs down and they rely on prompt deliveries; any delays to 
shipments will have an effect on their businesses and, therefore, shipping companies will have to 
consider this to remain competitive.

Preventative measures

The IMO began a project in order to raise awareness of the piracy problem in 1998, aimed at government 
representatives from countries that suffer from piracy to try to work towards solutions. The project also 
carried out a number of evaluation and assessment missions, with the aim of promoting the 
development of regional agreements on the operation of counter-piracy measures. Although projects 
like these can provide tangible results, their benefits may be limited when dealing with failed states and 
states that do not have the resources to tackle the piracy problems either on the land or at sea. Even 
without state interventions there are a number of physical measures that shipping companies and 
seafarers can undertake to reduce the vulnerability of attacks by pirates. If the gangs consider the vessel 
that they are planning to target is challenging this may prove to be a deterrence, but it is possible that a 
determined gang may work through the majority of countermeasures that are currently available if they 
are desperate enough. Providing countermeasures may reduce insurance premiums for vessels, but they 
are also likely to be expensive which may discourage their use. For example, the cost of employing a 
four-man PMSC team for a transit through the Gulf of Aden can be as much as $100,000. There are also 
some legal and practical aspects that must be considered before some deterrent measures can be 
employed and these are covered later on.

Barbed wire and razor wire is viewed as a relatively simple countermeasure that can be employed as it 
can restrict access to would-be attackers, but it is expensive, easily countered and has impracticalities and 
environmental issues. Many ports do not permit ships to have barbed or razor wire fitted when they dock 
which means a crew may elect to dispose of the installations overboard when they are through the risk 
area, or opt to have rapidly deployable and recoverable kits fitted that do comply. But all of these 
measures can be overcome with wire cutters and blankets and some pirate gangs take narcotics like khat 
to give them courage for the attack and to soften any pain from interaction with barbed wire or electric 
shocks.

Long Range Acoustic Devices (LRADs) and water cannons can be employed by vessels as a further means 
of deterrent and are relatively cheap to install. They can be used manually or remotely to deter attacking 
craft as they can make them unstable and make boarding a ship more difficult. The LRADs can have an 
effective acoustic deterrent range out to 300 metres and can transmit clear, directed, audible warnings 
out to 2000 metres. However, they can be overcome by pirates who use multiple skiffs to attack as a ship 
is unlikely to be equipped to deal with a multi-axis assault. A simpler method is for the pirates to just use 
ear protection.

By far the most imposing deterrent available is the presence of dedicated anti-piracy patrols. Some states 
are able to provide these themselves whilst others will rely on international efforts to support them. 
Nations such as the US, Japan, China, Russia, India and South Korea all provide stand-alone dedicated 
anti-piracy operations in piracy hotspots, and multinational operations like NATO’s Operation OCEAN 
SHIELD, the Combined Naval Task Force’s TF151 and the EU’s Operation ATALANTA currently provide 
dedicated naval anti-piracy missions around the Gulf of Aden. These operations, although effective, are 
constrained by resources both in terms of the number of vessels that are available to patrol vast areas, 
and the cost of providing them. In a 2010 report, the ICS estimated that $2 billion is spent each year on 
naval operations off the coast of Somalia alone. Therefore, although these are a highly effective 
countermeasure, their long term sustainability in terms of affordability should be questioned.

If vessels are to reduce the risk of attack, captains may elect to alter their speed or, if transiting through 
the Gulf of Aden, they may elect to join a group transit through the Internationally Recommended Transit 
Corridor (IRTC). Although both of these measures will provide risk mitigation, both are likely to incur 
additional costs for shipping companies. For example, container ships, with a normal average speed of 
twelve knots, are advised to transit the high risk area around the Gulf of Aden at speeds between 
eighteen and twenty knots. However, this brings additional costs. It is estimated that the cost of the
increases in speed for container ships alone in the high risk area in 2011 was approximately $2.7 billion. The increases in speed are based on advice from the Best Management Practice version 4 (BMP4) booklet produced by the IMO that highlights that, “to date, there have been no reported attacks where pirates have boarded a ship that has been proceeding at over eighteen knots.” Additionally, to join a group transit a vessel is likely to either have to wait for a slot time or to speed up to make the slot time, both of which will incur increased costs.

In addition to the costs to the maritime industry, local economies will also suffer as a result of piracy. Over fishing by foreign fleets and the risk of fishing in dangerous waters means that exports of fish products from piracy affected countries have declined by 28.5 percent since 2006. The dangers in the waters also have a knock-on effect to the tourist industry. East African countries such as Kenya, Mauritius and the Seychelles have noticed a significant drop in revenue as tourists stay away for fear of being caught up in the crime. Some of the West African countries that are affected by piracy find that it damages their economies, and the increase in insurance premiums for vessels going to West Africa has resulted in significantly reduced trade. In Benin, on the West African coast, half of the government’s income is derived from taxes on trade with 80% of this trade coming through the port of Cotonou. The drop in trade as a result of the increased risk to shipping led to a 28% loss in Benin’s government revenue. The second order effects of this piracy are felt by threats to the livelihoods of the country’s population as well as increases to the cost of imports and decreases to the competitiveness of exports. The threat of piracy around the Gulf of Aden will lead to some operators deciding to route their vessels around the Cape of Good Hope. It is estimated that this will result in a drop in revenues of around $642 million per year for Egypt from the decrease in use of the Suez Canal, which in 2008 was the third largest source of the country’s foreign currency income. It is likely that the threat of piracy and kidnap is also having a major effect on the tourism industry for littoral states. The Kenyan Tourist Board claimed in 2010 that, as a result of piracy, they saw a decrease of tourists from cruise ships into Mombasa from 11,000 to 500, a drop of 95%

One aspect that rarely makes the headlines in the counter-piracy debate is that of the potential environmental and ecological disaster that could occur as a result. There is the potential for a captured or abandoned ship either colliding with another ship, or against rocks, or being catastrophically damaged during an attack. These events could result in a massive oil spill similar to the Exxon Valdez in 1989 which cost an estimated $9.5 billion, or a spill of liquid gas or chemicals which could result in unimaginable and lasting damage. An event like this could set back the economies of fragile states many years, if not decades, and could also result in the closure of shipping lanes which would lead to diversions and further costs to the maritime industry. Although there is little that can be done to control a ship that is captured, this scenario is both low probability and high impact so should be considered when assessing the risks and costs of piracy.

The solutions

Having looked at a number of the causes and effects of piracy, this paper will now consider some of the solutions. Like any organised crime in the world today it is unlikely that piracy can ever be eradicated completely for two main reasons. The first is that for every countermeasure that is developed, for every patrol that is established and for every land-based initiative that is implemented, gangs that are intent on making money will endeavour to find an alternative or a work-around solution. The second reason is purely down to resources and cost benefit analysis. Although the current costly efforts by the international community to control piracy have made a significant impact, it would be both too expensive and physically impossible to police and patrol all of the oceans and, therefore, there will always remain an opportunity for pirates to carry out their crime.

So if piracy cannot be economically brought to an end, how can it best be managed? The IMB has produced comprehensive guidance to maritime operators for the prevention and suppression of piracy, and also produces the BMP4 booklet. This booklet provides specific guidance for ship operators and masters of ships transiting the high risk area near Somalia. It also allows maritime operators to understand the vulnerabilities of vessels, cargoes and valuables, how best to prevent boarding by pirates.
and what to do if a vessel is boarded. Following advice like this can be a major step forward in controlling and managing the piracy problem.

**State owned maritime patrols**

The role that international navies play in counter-piracy operations is crucial as, in the absence of credible coastguards and law-enforcement, they collectively provide good order at sea and can provide the only means of legitimate security enforcement. Although pirate gangs may provide a credible threat to a tanker when armed with automatic weapons and RPGs, they are no match for a well-equipped warship. It is argued that in the current Global War on Terror, the navies of the world have had little in the way of visibility in comparison to land or air forces and that counter-piracy operations allow them to take some of the limelight. This shift of focus sees them moving from the realist perspective of traditional maritime security to a liberalist’s approach of multilateral cooperation. Although this change of focus is evident, there can be little doubt that the security the navies are providing is beneficial to global trade. Additionally, the international cooperation between navies allows them to increase interoperability, expand situational awareness, strengthen relationships and to train together; opportunities that are not common in the current austere climate.

But can the investment in providing this level of security be justified? The EUNAVFOR had a budget of over eight million Euros in 2011 just for patrolling the 3.2 million square miles around Somalia. In a research report from 2009, Hansen detailed the cost of deploying one Norwegian frigate to the Gulf of Aden for six months as being roughly the equivalent of paying for 100,000 Puntland police officers for the same period. It is widely recognised that piracy cannot be tackled from the sea alone, so perhaps it is time for the international community to re-focus its resources on where they can be most effective and tackle the root causes of the problem, not just the result.

One of the economic drawbacks of the increased naval presence, and the subsequent decline in piracy within the territorial waters, is that there will be a reduction in the pirates’ unofficial policing ‘role’. The presence of pirates has led to a significant decrease in illegal fishing, but as the gangs are deterred or move further away from coastal waters this role will no longer be filled as it is not one of the functions of the international naval forces. Local fishermen in Indonesia, Somalia and Kenya have celebrated the deterrent effect that the pirates have had on scaring away commercial fishermen, as it has allowed fishing stocks to recover which allows them to increase their catches. Although this by-product of piracy is beneficial to some areas of littoral states, improvements in governance, coastguard and security forces should see that this issue is formally managed effectively in the future.

In addition to providing naval assistance and patrols, it is important for the international community to try and develop an indigenous capability to tackle piracy, and to develop a secure environment at sea. By sponsoring and training credible coastguard and law enforcement operations, the international community will be able to focus its resources away from maritime patrolling, and on to some of the many other projects that can lead to longer term stability. After international training assistance, Somalia graduated its first 500 coastguard recruits in September 2009 and in July 2013 the Dutch Atlantic Marine and Offshore Group signed a contract with the Federal Government of Somalia for the development of the Somali Coastguard, including the provision of equipment, training and management services. Although these appear to be very positive steps, the international contribution has been met with resistance from the Somaliland Ministry of Fisheries and Marine Resources, who see the Dutch company’s involvement as invasive and have warned them to stay out of Somaliland’s territorial waters as Somaliland “does not need any foreign company to do its duties.” From this reaction it is clear that, in addition to financial resourcing, a comprehensive approach that includes strategic messaging is required to ensure all actors are working together to achieve the same aim.

**Ship-based countermeasures**

Some of the basic ship-based countermeasures that are recommended in the BMP were touched upon earlier in this paper. However, one of the most widely discussed and debated measures currently being both considered and used, is the employment of PMSCs on board ships. Arming crews or the
employment of armed PMSCs are often cited as a potential solution to combat piracy and, in 2012, it was estimated that PMSCs were used on around 40% of large vessels in the high risk area. As of May 2013, no vessel carrying armed PMSC personnel had been successfully captured by pirates. However, the introduction of armed crews or PMSCs into the maritime industry brings with it a significant number of issues which need to be carefully considered. The quality and quantity of the weapons that pirates use has changed over the past decade and has moved from machetes, knives, clubs and pistols and now includes automatic rifles and RPGs. Although the weapons are rarely used against crews during an attack, the introduction of armed crews or PMSCs has the potential to make pirate gangs more hostile and to use more force. This posture could lead to an escalation in the armament held on both sides and it is likely to be a “race that merchant sailors cannot win.” This would, therefore, mean that arming personnel on board ships could potentially make them less safe rather than more safe.

Some countries may not permit armed crews or PMSCs to enter their national waters or carry weapons on board their ships, and these are being monitored by insurance companies and the ICS. Many ports will also not permit vessels to be docked with armed crews or PMSCs. If permitted, they may demand a company to prove that its weapons have been bought, stored, and exported in accordance with all applicable national regulations, which would then cause administrative problems for shipping companies. The weapons issue can mean that some PMSCs may throw their weapons overboard before entering ports or territorial waters which increases the costs of their operations and is not environmentally or economically sound.

The issue of multinational crews on board commercial vessels of a different flag state transiting another nation’s territorial waters becomes extremely complex and faces panoply of legislative issues. Many questions can arise from the issue including who would authorise the drawing of weapons from their secure area? What would happen if a pirate was taken prisoner in an armed struggle? How far could an armed individual go in the use of force, would it be limited to self-defence, defence of another or defence of property?

Although there are a number of legal issues to the armament of crews which are far from being resolved, the cost-benefit analysis is likely to be a key factor in deciding whether crews should be armed or armed PMSCs carried on board. With many insurance companies offering significant discounts for vessels which carry armed security personnel, the risk of not losing crews, cargo or vessels to pirates or having vessels offline whilst ransom negotiations are underway all give a powerful incentive for shipping companies to employ PMSCs. Interim guidance has been issued by the IMO to help shipping operators come to a decision as well as by governments, including the UK, but this still leaves the complicated legal decisions with the shipping companies and ship’s captains. Steps are underway to try and regulate the use of PMSCs, the Security Association for the Maritime Industry has made some progress with background checks on PMSC credentials and a quality assurance framework scheme was released in late 2012. The aim of this scheme is to try and address concerns with legal issues and safety concerns for seafarers operating under varying national regulations. Steven Jones, director of the Security Association for the Maritime Industry said the last thing that anyone wants is a “Blackwater of the sea.”

Although these regulatory measures are a step in the right direction to help prevent the exploitation of the security market by unregulated private security companies as happened in Iraq, to be truly effective they still need the buy-in from all flag states and for them to clearly specify their position so a clear legal policy can be determined.

**Transnational organisations**

In addition to preventative measures by maritime operators, coordination of resources amongst littoral states effected by piracy and by organisations such as the EU, NATO, and the UN is essential to ensure understanding is shared, lessons are learned and the maritime industry is kept up to date with the latest threats and developments. During the boom years of piracy in the Straits of Malacca, Indonesia, Singapore and Malaysia all signed the Operation MALSINDO agreement to coordinate their anti-piracy efforts and this led to a sharp and sustained decline in the number of attacks. The establishment of organisations such as the Maritime Security Centre Horn of Africa and the Regional Centre for Maritime
Security in Central Africa, in addition to the multinational coordination centres in the region, are critical for this understanding. A Contact Group on Piracy off the Coast of Somalia (CGPCS) has also been established and is a voluntary international forum created in January 2009, in response to UN Security Council Resolution 1851. The CGPCS has been a key body in leading the international community’s response to Somali piracy, and has a role of coordinating international actions and facilitating dialogue to counter piracy off the coast of Somalia. The group has achieved significant progress in countering piracy at sea as well as carrying out actions to “coordinate international efforts to disrupt the financial networks of pirate leaders and their financiers.” These transnational organisations are able to use mass and breadth to tackle piracy in a comprehensive manner by sharing information, pooling resources and providing the manpower and equipment to both understand the problem and to continue to try to tackle it.

**Confronting piracy at its source**

Pirates need safe harbours and bases to operate from and land-based support to provide them with the supplies and resources they need to operate. Therefore, synchronising counter-piracy operations with land-based activity will help to target the root causes of piracy, and should be a key strategy to tackle the problem. However, the difficulty comes when piracy originates from weak, fragile or failed states as it could be unlawful for an international force to enter a sovereign territory, or it could be too dangerous. Therefore, interventions to target the causes of piracy must be multifaceted and include support to stabilise or develop governance, and development of local law enforcement. Peter Chalk considers that it would not require a large military input as, in general, local communities already support reforms and could provide local knowledge and intelligence. NATO has taken steps to engage with key Somali tribal leaders in order to counter piracy and, in 2010, the EU launched a training mission to develop the Somali security services with the intention of recruiting up to 2000 individuals.

The improvement of infrastructure to support developing economies will lead to better economic and employment opportunities, so individuals have better prospects of legitimate employment and do not have to turn to crime as a means of providing subsistence. Cutting the links to pirate financiers is on the UN’s agenda and is called for under resolution 2020, which recognises the “need to investigate and prosecute those who incite or intentionally facilitate piracy operations” successfully achieving this, larger pirate gangs will lose the ability to conduct their operations, although smaller scale operations will still be possible. Similarly, the financing of the multi-million dollar stolen oil industry in Nigeria needs multinational action in order to close off markets, as it is too big a problem for the weak government of Nigeria to deal with alone.

**Deterrence and legal processes against piracy**

To be a credible deterrent, the legal processes that can be used against pirates must be robust and this relies on effective state governance and, to a degree, international cooperation. It has been documented that the levels of arrest for pirates can be as low as 1% and lengths of sentences can vary from two to twenty-four years and average ten years. States may be reluctant, at times, to carry the legal process through to conclusion due to the cost and logistics of doing so, however, a number of high profile cases show that this is not always the case. For example, in New York in February 2011, Abduwali Muse pleaded guilty to hijacking the American-flagged ship *Maersk Alabama* and kidnapping its captain and was sentenced to thirty-three years imprisonment. Although the US government decided to drop a charge of piracy, which carries a mandatory life sentence, the judge imposed a sentence at the top of the range as she saw deterrence as critical factor in the case.

To tackle the piracy problems in the most effective way, the legal processes should be aimed at the transnational criminal organisations that finance and organise piracy, not just at the foot soldiers who are involved in the operations. Removing the finance and leadership will prove more effective than targeting the foot soldiers who can easily be replaced by willing individuals who see an opportunity to get out of poverty. This action is endorsed for Somalia, under UN resolution 2020, which calls for nations to establish specialized anti-piracy courts to investigate and prosecute key figures of criminal networks involved in piracy, as well as the pirates themselves. However, the international community must take
its share of the responsibility to prevent countries charged with the judicial process from becoming overwhelmed. In April 2010, the Kenyan government declared it was at capacity and would not take on any more piracy cases, accusing the international community of failing to share the burden of responsibility and cost of legal processes. Therefore, to allow deterrence to be an effective measure against potential pirates, significantly more investment, such as that already provided by the international community to open a fast track court and holding facility, must be made.\(^{103}\)

This investment will support the prosecution and detention of pirates and other criminals and help to actively target pirate leadership which can lead to the break-up of the criminal networks. As well as removing the leaders’ ability to organise more attacks, their arrest should serve as a potential deterrent and show the leaders that they are being targeted. Mohamed Abdi Hassan was detained in Belgium for belonging to a criminal organisation, kidnapping and having allegedly organised the 2009 hijacking of the Belgian vessel *Pompei* as well as the hijacking of several of other commercial ships. Allegedly Hassan was one of the most influential pirate leaders from Somalia. However, at the time this was only the first prosecution of a pirate leader by the international community,\(^{104}\) showing that a change in focus is required in order to dismantle the pirate networks.

The ICS estimated that in 2010 the cost of piracy trials and imprisonment was around $31 million\(^{105}\) but this is nothing compared to the cost of ransoms, or additional insurance premiums paid by shipping companies. Under UNCLOS article 105, whilst on the high seas or outside of sovereign waters any nation can arrest persons suspected of piracy, and the courts of the state which carried out the seizure may decide upon the penalties to be imposed.\(^{106}\) Although, in principle, this gives a good deal of freedom to prosecute, in practice, it also reveals a number of practical difficulties. If an international patrol was to arrest a pirate, then under article 105 it could lawfully try them in its own courts. However, the holding of the prisoners at sea is impractical as ships are not designed for this purpose and extended detentions could be considered against the human rights of a prisoner.\(^{107}\) Furthermore, the logistics of transferring them for trial are likely to be both difficult and expensive, and some states who undertake anti-piracy patrols are reluctant to pay the costs associated with trials due to their uncertain outcomes.\(^{108}\) This has led to captured pirates being disarmed and then released and it is reported that between January and June 2010, the commanders of EU and NATO naval forces released around 700 suspected pirates.\(^{109}\) Once the suspects have arrived in the arresting state’s country, or on completion of a sentence, they may wish to try and claim asylum, as recently happened in the Netherlands, and this is unlikely to be in the arresting nation’s interests.\(^{110}\) Even after release from prison it may be that a country feels it is too dangerous to deport an individual, and there are concerns that “trials in European courts would encourage, rather than deter, pirates from committing crimes of piracy.”\(^{111}\) To overcome this problem, the US, UK and the EU have established agreements with Kenya, the Seychelles and Mauritius to allow extrajudicial transfers of captured pirates for criminal investigations and prosecutions.\(^{112}\) However, it appears that the appropriate investment has not been made to deal with those who are caught. Kenya’s court infrastructure is deemed to be inadequate to deal with the demands of Somali piracy and in 2010 it had a backlog of approximately 800,000 cases.\(^{113}\)

The Djibouti Code of Conduct, which was signed on 29 January 2009 by all of the East African, Red Sea and Gulf of Aden states, in principle goes a long way towards addressing the cooperation and resourcing concerns with the investigation, arrest and prosecution of persons who are suspected of having committed acts of piracy.\(^{114}\) In a similar fashion, the IMO and the Maritime Organization of West and Central Africa developed a Memorandum of Understanding (MOU) in July 2008. This MOU was signed by fifteen states to establish a coastguard network in West and Central Africa to assist in the policing of the waters and to help combat piracy operations. This was then followed by the formal acceptance of a code of conduct by twenty-two states, which incorporates elements of the Djibouti Code of Conduct, to further develop the coastguard capabilities of the West African states.\(^{115}\) This means that for the vast majority of the littoral states of Africa where piracy is an issue, a formal funding framework is now in place that can be used to tackle piracy; with the right leadership to take these agreements forward significant progress is achievable. If they prove to be successful, these codes of conduct could be used as a model for other pirate hotspots, but the rate of success is likely to be dependent upon the international contributions received into the trust funds in order to finance the operations.
Political and religious will and community support

One of the most successful actions to have reduced the instances of piracy in Somalia was in autumn 2006 when the Supreme Council of Islamic Courts (SCIC) was formed. The SCIC elected to put an end to piracy and announced that piracy was against the teachings of Islam. After launching a full-out offensive against some of the major pirate strongholds, piracy rates dramatically declined. Although this phenomenon was short-lived, as the SCIC was driven out of Mogadishu in 2007 by Ethiopian forces, it did show how much of a factor religion played in that part of the world. This example shows that this is an opportunity that can be utilised and that the employment of religious leaders in the future management of pirate gangs is one which could be considered.

Although the lure of money in areas of poverty will inevitably be one of the strongest influences for an individual to become involved in piracy, money alone cannot support a pirate. There also needs to an element of legitimacy for the pirates to be supported by their local communities, in order for them to be provided with supplies, food and shelter. This legitimacy can be provided through a belief that the pirates are acting in the interests of the communities by policing the waters to prevent illegal resource exploitation or pollution. However, this belief is countered by the anti-piracy narrative that is put forward by the international community and this is an area which can be developed with communities in order for them to alienate pirates and withdraw their support. Through targeted messaging it will be possible to influence communities who can then withdraw their support from pirate gangs. However, the maintenance of the counter-piracy narrative can only be sustained if communities feel the benefit in areas such as aid, employment and economic development. There must be a clear message given that wellbeing can be improved through cooperation and lawful behaviour.

The commercial benefits of piracy

Although piracy can be costly in terms of maritime patrols, ransom payments, injuries, deaths and increased commodity prices, there are now a number of industries which thrive and rely on its development and sustainment. Private security companies can “smell blood in the water and cash-in on anti-piracy operations.” With only an estimated 2% of the cost of piracy going to ransoms, the majority of the rest is returned to the economic circuits of those involved. Industries have grown significantly as the threat of piracy has grown, and many would cease to exist if piracy was to be eradicated, or if the threat from piracy was deemed to be negligible. Therefore, this paper argues that the current equilibrium between anti-piracy measures adopted by states and the maritime industry, and the actions of the piracy gangs, creates a symbiotic relationship with maritime insurers, PMSCs and law firms who conduct negotiations and arrange for payments. Although there is no evidence to suggest that any of these businesses openly support piracy, without it, many would not be able to operate on the scale they do.

Based on figures provided in the Oceans Beyond Piracy Report, the commercial value of the counter-piracy industry in the Gulf of Aden alone can be estimated as $363 million - $2.5 billion for the cost of security equipment and guards, and $460 million - $3.2 billion for piracy insurance (of a worldwide marine insurance market of $31.9 billion).

These figures do not include the fees for law firms who will be involved in negotiating and arranging payments and, it can be assumed, that they will not be inconsequential. The firms undoubtedly undertake a necessary role in the complex legal and international environment that the negotiations take place in, but they also stand to make significant profits from doing so. Law firms can run 24 hour hotlines and can be contacted and involved in the process within thirty minutes of a hijacking. They then become involved in matters, such as advising on insurance disputes over what is covered and what isn’t, conducting negotiations regarding the amount of ransom to be paid, arranging for the actual delivery of the ransoms and liaising with the families of crew members about possible further action.

The maritime insurance industry is likely to be one of the biggest beneficiaries of piracy. As long as piracy exists they will be able to charge a premium and make a profit. As the risks increase and decrease then premiums will vary, but with the waters off the coast of Somalia being deemed to be a war zone by the Joint War Committee at Lloyds, this will remain an expensive area in which to obtain insurance. It can be argued that insurance companies do not demand that ship owners take security precautions as it
would make pirate attacks more difficult. According to a publication from the German Institute for Economic Research "you cannot expect insurance companies to saw off the branch that is nourishing this insurance market." Whilst most ship owners and hull underwriters were understood to have had a difficult year financially in 2012, the 2013 report by the Shipbrokers BRS said that counter-piracy insurance had remained profitable showing that there is money to be made from piracy and that insurance companies can adapt.

PMSCs have seen a rapid growth as piracy has evolved, in a similar way to the development of security companies during the boom years in Iraq in the second Gulf War. They not only provide on-board personnel to deter pirates, they also provide escort vessels, consulting and intelligence, training, risk assessments and a number of other lucrative services. This rapid growth has brought with it high revenues and large profits. A report to the UK parliament in 2011 estimates revenues from private naval protection companies to be in the region of $680 million each year. This rapid growth raises concerns about the capabilities and effectiveness of some of these businesses which has led to attempts to regulate the industry. In 2012 it was reported that there were approximately 200 to 300 companies operating that offered private maritime security services, at least half of which are based in the UK.

**Conclusion**

In conclusion, this paper has shown that there are a significant number of interested parties in the ‘business of piracy’. These actors include the states who are trying to stop it, shipping companies and their employees who have to deal with it, legitimate security and insurance companies who profit from it and the pirate networks who are ultimately responsible for the crime and benefit directly from the goods they steal and the money they ransom. Throughout the chain there are secondary beneficiaries and also parties that are detrimentally affected. The beneficiaries include local communities who may profit from pirate wealth and corrupt government, law enforcement, military and industry officials who receive bribes and profit shares. Those that are indirectly disadvantaged include tourist industries, goods sellers who have to pay for higher transportation costs and, ultimately, tax payers who pay for the anti-piracy operations through taxation. There are also unmeasurable human costs in terms of lost trade, the psychological impact on those who have been caught up in piracy and those that work in fear of attack in the maritime industry.

The paper has shown that, although the nature of piracy can vary from region to region in terms of frequency, motive and method, the techniques to try to counter it can be broadly similar and rely on a combination of effective self-protection and deterrence measures, policing and patrolling by state maritime organisations and by addressing governance, economic and security issues from countries where piracy originates. In a similar way that insurance companies offer discounts for adding protection measures to a property, maritime insurance premiums must reflect piracy countermeasures and deterrent measures in order to help to reduce piracy further. Shipping companies must be given greater financial incentives for the employment of basic deterrent measures in accordance with guidance such as BMP4 and for the posting of effective anti-piracy watches. Additionally it has been suggested that vessel owners, crews and flag states should be potentially held responsible if they do not abide by BMPs to incentivise them to take at least basic precautions. “Getting to zero will be a very, very difficult thing to do” but managing the situation and getting to an acceptable and tolerable level is more easily achieved.

The long-term solutions to managing piracy depend on the creation and maintenance of a stable economic environment which is protected against the undermining and destabilising effects of corruption, an environment which is unfortunately ingrained in some parts of society and one that will be hard to change. The policing of pirate affected waters is very expensive and cannot be sustained indefinitely and, therefore, solutions must be found. Strong, determined governments that support economic growth and support a stable and reliable police force and coastguard are likely to be the most effective tools to guarantee success. These can be combined with powerful influence from religious and community leaders who can withdraw support from pirate gangs in order to isolate them. These principles are supported by the UN who call for a “comprehensive response by the international community to repress piracy and armed robbery at sea and tackle its underlying causes.” By
combining these approaches, states can return to protecting and fishing their valuable marine resources or, in the case of Nigeria, they can return to a legitimate trade in oil which will benefit all of the country, rather than just the corrupt few who benefit from bunkering and theft.

In addition to measures at sea, a consolidated land-based effort is critical in the fight against piracy and investment, resourcing and cooperation are key to its success. A recent strike by the EU NAVFOR on targets near the pirate haven of Haradheere destroyed fuel, outboards and speedboats. The intention of the attack was to disrupt logistics on land and send a strong message to the pirate gangs and their financiers to show them that they cannot hide ashore. However, unless operations like this becomes a regular weapon in the international communities’ armoury, it is unlikely to become a significant deterrent and gangs will still continue to operate with relative impunity.

As part of the overall solution, the international community needs to consider the balance of effort between the resource put into counter-piracy around the Gulf of Aden and those committed to the Gulf of Guinea, which is far less exposed in the international forum. There is a significant disparity in resource commitments and guidance available which will ultimately mean that the problem around the West coast of Africa is not contained and will flourish as a consequence. Some Western navies do occasionally patrol the area and provide capacity building training, but it is not enough. This capacity building should be seen as a “spend to save” measure with consideration given to the long-term economic prosperity of the region, as well as the harmful effects that piracy causes to the maritime industry.

The current statistics provided by the IMB show that although piracy remains an issue, it is on the decline and this is attributed to “the combined efforts of the Navies…along with the increased hardening of vessels and BMP4 compliance, employment of Privately Contracted Armed Security Personnel and the stabilizing factor of the central government within Somalia.” When the falling rates of piracy are compared against the vast volume of shipping that still transit these waters, and is still on the increase, the value of committing any more resources to the fight can be questioned. The statistical probability of a successful attack around the Gulf of Aden in 2013 was less than 0.5% and the value of losses from piracy at sea are comparatively insignificant when compared to the losses of cargo on land. However wide the range of estimates are for the global cost of piracy they are extremely small in comparison to the value of global trade that is transported by sea, which in 2011 amounted to approximately $13.4 trillion. These figures, despite being an approximate, show that whilst piracy is an issue, it must be considered in the perspective of its actual cost compared to the cost of trying to manage it. This paper concludes that piracy on the high seas can only be reduced further, but will never come to an end. The only way for this reduction to happen effectively will be by the prioritization of finite resources so that more emphasis is made on the development of weak states, not just on the policing of their criminals.

Endnotes

1 IMO, “Introduction to IMO,” http://www.imo.org/About/Pages/Default.aspx
10 The Oxford Dictionary of English (3 ed.) defines a Failed State as one whose political or economic system has become so weak that the government is no longer in control.
13 Yikona. Pirate Trails. 28.
14 Ibid, 46.
63 The high risk area is defined by the IMB as stretching from the Suez and the Strait of Hormuz to the North, Sri Lanka to the East and Madagascar to the South.
66 NATO Shipping Centre, “Group Transits Through the IRTC,” para. 2. http://www.shipping.nato.int/operations/OS/Pages/GroupTransit.aspx
67 Yikona, Pirate Trails. 33.
69 Bowden, The Economic Costs of Maritime Piracy, 23.
79 Somaliland and the federal government of Somalia and the Atlantic Marine and Offshore Group deal does not concern us-Director-General of Fisheries.
85 The flag state of a commercial vessel is the state under whose laws the vessel is registered or licensed
86 Instituto Affari Internazionali, The Threat of Contemporary Piracy, ix.
91 Lloyd’s Market Association, “Joint War Committee,” para. 5.
92 Ong-Webb, Piracy, Maritime Terrorism and Securing the Malacca Straits. 59.
94 Yikona, Pirate Trails. 33, 34.
95 Chalk, “Piracy off the Horn of Africa,” 68.
98 Katsouris, Nigeria’s Criminal Crude: International Options to Combat the Export of Stolen Oil, ix.
Courage: Cardinal virtue or core value?

Rev’d Alan Steele MBE MPhil

Introduction

Courage is one of the British Army’s core values, so, when I became an Army Chaplain, I naturally read *The Anatomy of Courage*, a study of fortitude written by Lord Moran on the basis of his experience as a Medical Officer on the Western Front during the First World War. One of Lord Moran’s conclusions struck me forcibly:

“a man of character in peace becomes a man of courage in war…… Man’s fate in battle is worked out before war begins.” (Moran 1945:170)

Over the years, on operational deployments in various parts of the world, as I have lived and worked with soldiers, sometimes under the stress of intense combat, particularly in Afghanistan, the truth of this conclusion has been repeatedly brought home to me: “Man’s fate in battle is worked out before war begins.” I must, therefore, start with an apology because my research methods have been unscientific; I have started with a conclusion, which experience has led me to believe is true, and have only then sought the evidence to explain it. This paper is an interim report, if you like, describing how far I have got in my attempts to understand it.

First, however, I would briefly like to discuss the various terms associated with the idea of courage. I am aware of the very varied etymological roots of these terms: bravery and gallantry originally denoting ‘ostentatious behaviour’ and ‘gay display’; while fortitude and valour are ultimately derived from words meaning ‘strength’; heroism comes from the Greek name for those outstanding individuals, leaders of men, who won immortal renown both as “speakers of words and doers of deeds”, as Homer puts it; and courage itself originally meant an enlivened and emboldened heart, so Chaucer can describe the invigorating effect of spring upon birds:

“small fowls make melody, that sleep all the night with open e’e, nature so stimulates them in their hearts”2 – “in hir corages.”

No doubt in an heroic age and, indeed, in a chivalrous one, there was the need for such distinct terms, but in contemporary usage I believe that all these terms have effectively become synonymous and that, nowadays, it means much the same thing whether someone is described as being courageous, brave, valiant, daring, or heroic. Indeed, this blurring of vocabulary and the consequent loss of our ability in common parlance to make, for example, a clear, succinct distinction between the man who has true bravery and the one who is all bravado, is part of what, in this paper, I describe as the devaluation of courage.

Courage in the First World War

Sir Winston Churchill, in an essay published in 1937, stated,

“Courage is rightly esteemed the first of human qualities, because… it is the quality which guarantees all others.” (Churchill 1947b:167).

Churchill could speak from wide military experience, both as a cavalry officer in the small wars of Empire and also as a battalion commander on the Western Front in 1915-1916. Churchill’s belief that courage was the premier human quality was shared by others who served during the First World War. Lt Col A.A. Hanbury-Sparrow wrote of his officers:

“Were they or were they not braver? That was your criterion… For the act of being brave compelled the utilisation of the whole reserve of moral force that lay in a man.” (Keegan 2004:275).

Not just officers, but chaplains also maintained the primacy of courage. For example, an Australian chaplain, Rev’d K. Anderson, wrote:
The influence of a chaplain over the men depends on one thing – his obvious physical bravery. Everyone can value courage, for all know the meaning of fear. All things may be forgiven to the chaplain who shows himself prepared to share their dangers; nothing can mitigate the failure of the man who is not.” (Smyth 1968:166).

Another example of this attitude is the celebrated advice given by Rev’d G.A. Studdert Kennedy to Rev’d Theodore Hardy, when the latter first arrived in France as a Chaplain in 1916:

“Live with the men, go everywhere they go. Make up your mind you will share all their risks, and more, if you can do any good. The [front] line is the key to the whole business. Work in the very front and they will listen to you; but if you stay behind, you’re wasting your time. Men will forgive you anything but lack of courage and devotion.” (Smyth 1968:175).

While the men who served in the First World War might question whether it truly was “sweet and seemly to die for one’s homeland” and lament the waste of war, not even the most disillusioned post-war writers ever questioned the value of courage. If anything, the experience of the war, an experience through which almost the entire manhood of the nation passed, reinforced the belief in the primacy of courage. Lord Moran, reflecting during the Second World War on his experiences in the First, wrote:

“We came to exalt courage. Was this man staunch in battle? That was the acid test to which every soldier in France was put. And now because courage is rare, because it alone stands between us and the ruin of our cause, we must once more acknowledge and confess the primacy of courage.”

Bizarrely, J.M. Barrie, who had no military service, chose ‘Courage’ as the topic for his 1922 Rectorial Address to St Andrew’s University, delivered, as he acknowledges, to men, many of whom had seen active service in the war. Here is his conclusion:

“Courage is the thing. All goes if courage goes. What says our glorious Johnson of courage: ‘Unless a man has that virtue he has no security for preserving any other.’ We should thank our Creator three times daily for courage instead of for our bread, which, if we work, is surely the one thing we have a right to claim of Him. This courage is a proof of our immortality.” (Barrie 1922:37)

I shan’t vex you with further quotation. It was a presumptuous theme to speak on before such an audience, and, I suspect, probably infuriated those in it who had seen war service, but it accurately reflects, I believe, the pre-eminence to which courage was raised as the supreme value of post-war British society.

The traditional western view of courage

How different then, to turn from this florid exaltation of courage to the sober statement by St Ambrose, written almost 17 centuries before:

“Courage without justice is a means of evil.”

Nine hundred years later, quoting Ambrose, St Thomas Aquinas affirmed:

“A man does not expose himself to mortal dangers except in order to preserve justice. Therefore the praise of courage to a certain extent depends on justice.”

It is worthwhile remembering that this is not a peculiarly Christian ideal. In these passages, both Ambrose and Aquinas are discussing the Four Cardinal Virtues; virtues derived from the pagan Greeks and, hence, virtues which are accessible to anyone, whether or not they are Christian. In the Fourth Century BC, Aristotle, for example, declared:

“It is for a noble end that the brave man endures and acts as courage directs.”

Cicero, too, in the First Century BC, states of courage:
"But if the elation of soul, which is shown in danger and toil, is devoid of justice and fights not for the common weal, but for its own advantage, it is a vice; and it not only has no measure of virtue, but, rather, is monstrous and repulsive to all humaneness."

The most complete discussion of courage understood as a virtue, is that by Aquinas in the Summa Theologiae, in which he synthesizes the previous discussions of the Classical philosophers, particularly Aristotle, with those of the early Christian Fathers. He specifically asks the question whether courage excels all other virtues (op. cit. 2a2ae. 123, 12), and answers firmly in the negative. While courage is one of the Four Cardinal Virtues, it is not the foremost virtue and is subordinate, not only to justice, but also to prudence. Josef Pieper, the German Thomist philosopher, summarised Aquinas's argument as follows:

"Prudence and justice precede fortitude. And that means, categorically: without prudence, without justice, there is no fortitude; only he who is just and prudent can also be brave; to be really brave is quite impossible without the same time being prudent and just also. Under the direction of prudence, the good of man becomes compellingly evident. Justice primarily brings about the actual realisation of this good. Fortitude, therefore, by itself, is not the primary realisation of the good. But fortitude protects this realisation or clears the road for it." (Pieper 1966:123-125)

In essence, the traditional Western moral philosophy of the ancient Greek and Roman philosophers and of the medieval scholastic theologians is predicated on the concept of virtue: the characteristic excellence of being a complete, integrated human being. According to this moral philosophy, a man, by habitual practice, will both nurture and form a character of excellence by which he will be able to respond in a morally appropriate way in whatever circumstance he finds himself. The virtuous man's character is founded on the four Cardinal Virtues: Prudence (prudentia) entails understanding the objective reality of the world, not an understanding based on personal sentiment, opinion, or ideals; it involves relating to the world as it actually is. Justice (iustitia) entails ensuring that one's own conduct is right in relation to the wider community, it is based on the acknowledgement that others, too, have claims and rights, which must be guaranteed. As the ancient Roman Law Codes describe it:

"Justice is the firm and enduring resolution to render to everyone his due."8

Courage (fortitudo) entails being willing to face social and bodily discomfort, maiming or death to attain the good, the just, the true. Temperance (temperantia) entails self-control and self-discipline in one's lifestyle, actions and thoughts; not giving in to self-gratification or self-indulgence, but being moderate in one's enjoyment of legitimate goods, so that one is not destructive of either oneself or society as a whole (Pieper 1966; 1991 & Lewis 1943).

This raises the perennial question, first asked by Socrates, of whether or not all virtue is one. It is clear that the four Cardinal Virtues, at any rate, are very intimately connected with each other. For example, fortitude is fundamental to the practise of virtue, as C.S. Lewis noted in a BBC radio broadcast during the Second World War,

"Fortitude includes both kinds of courage – the kind that faces danger as well as the kind that 'sticks it' under pain. 'Guts' is perhaps the nearest modern English. You will notice, of course, that you cannot practise any of the other virtues very long without bringing this one into play." (Lewis 1943:14).

In saying this, Lewis is merely echoing what Aristotle, Ambrose and Aquinas had said before him. Nonetheless, Pieper's statement that "only he who is just and prudent can also be brave", is equally an accurate reflection of traditional Western moral philosophy. Indeed, Aquinas asserts that "all moral virtue is necessarily prudent,"9 because, in any given situation, one cannot be appropriately just or brave or temperate, if one does not understand the objective reality of the situation through prudence. Gilbert Meilaender discusses this question of the unity of virtue and concludes with "the fundamental paradox" that:

"One must do what prudence requires in order to be virtuous. But one must be virtuous in order to do what prudence requires. Put more simply: doing what is right requires being good, but we can become good only by doing what is right." (Meilaender 1984:26).

It is difficult for contemporary people to resolve this paradox of traditional virtue ethics, because – to speak very simplistically - contemporary ethical theories focus on how we ought to act (deontological ethics) or the outcomes
of our actions (teleological ethics), while virtue ethics focuses on the formation of outstanding moral character. In other words, contemporary moral philosophy revolves around discussions of human actions and how to make reality conform to subjective human values, while virtue ethics revolves around how to train the human soul to conform to objective reality. Without prudence to discern objective reality such conformation of the soul is impossible, while the training of the soul by the habitual practice of virtue, nourishes and develops prudence.

“There is a difference between doing some particular just or temperate action and being a just or temperate man. Someone who is not a good tennis player may now and then make a good shot. What you mean by a good player is the man whose eye and muscles and nerves have been so trained by making innumerable good shots that they can now be relied on. They have a certain tone or quality which is there even when he is not playing… In the same way a man who perseveres in doing just actions gets in the end a certain quality of character. Now it is the quality rather than the particular actions which we mean when we talk of the ‘virtue’ of justice.” (Lewis 1943:14-15)

The demoralisation of western society

Sir Winston Churchill, in his autobiography, My Early Life, published in 1930, wrote:

“I wonder often whether any other generation has seen such astounding revolutions of data and values as those through which we have lived. Scarcely anything material or established which I was brought up to believe was permanent and vital, has lasted. Everything I was sure or taught to be sure was impossible, has happened.” (1947a:67)

Later in the book, discussing the change (for the worse) in habits of alcohol consumption amongst Army officers, he wryly observes,

“Of course we all know much better now, having been civilised and ennobled by the Great War.” (Churchill 1947a:126)

Churchill is not alone in remarking on this revolutionary divide between the 1800s and the 1900s. C.S. Lewis, in his 1954 inaugural lecture as Professor of Medieval and Renaissance Literature at Cambridge, discussed cultural epochs in Western history and noted that:

“the greatest of all divisions in the history of the West [is] that which divides the present from, say, the age of Jane Austen and Scott.” (Lewis 1962:17)

Gertrude Himmelfarb has explored this seismic change in our moral understanding in her book The De-Moralization of Society: From Victorian Virtues to Modern Values – the title of which pretty much sums up her argument. She noted that the radical change in moral language from ‘virtues’ to ‘values’, was consciously initiated in Germany in the 1880s by Friedrich Nietzsche as a deliberate revolution in moral philosophy, but that in English usage it entered both philosophical and colloquial language almost imperceptibly. The actual effect of the change in moral terminology, however, was the complete transformation of both moral philosophy and the common understanding of morality. The word

“‘Values’ brought with it the assumptions that all moral ideas are subjective and relative, that they are mere customs and conventions, that they have a purely instrumental, utilitarian purpose, and that they are peculiar to specific individuals and societies…… Values, as we now understand the word, do not have to be virtues; they can be beliefs, opinions, attitudes, feelings, habits, conventions, preferences, prejudices, even idiosyncrasies – whatever any individual, group, or society happens to value, at any time, for any reason. One cannot say of virtues, as one can of values, that anyone’s virtues are as good as anyone else’s, or that everyone has a right to his own virtues.” (Himmelfarb 1994:11-12)

A consequence of this fundamental revolution in moral language and ethical understanding is that the traditional Western concept of virtue appears ludicrous to contemporary people. This is particularly true of the assertion that courage is necessarily prudent, as Pieper (1966:4-5) noted:
“To the contemporary mind, prudence seems less a prerequisite to goodness than an evasion of it. The statement that it is prudence which makes an action good strikes us as well-nigh ridiculous...... In colloquial use, prudence always carries the connotation of timorous, small-minded self-preservation, of a rather selfish concern about oneself...... Certainly the common mind regards prudence and fortitude as virtually contradictory ideas...... Modern man cannot conceive of a good act which might not be imprudent, nor of a bad act which might not be prudent. He will often call lies and cowardice prudent, truthfulness and courageous sacrifice imprudent.”

The overthrow of the language of virtue, seen in the debased connotations given to the words ‘virtue’, ‘prudence’, and ‘temperance’ in everyday speech, is also reflected in the contemporary use of words relating to courage, such as ‘hero’, ‘fortitude’, ‘bravery’, and ‘courage’ itself. One example of this is the ‘sentimentalisation’ of courage seen in the contemporary media, in which, for example, anyone who has had military service is described as a ‘hero’, so that the term now almost means the same thing as ‘veteran’. Interestingly, this sentimentalisation also occurred after the First World War, so a 1918 newspaper headline in the United States reads “D.C. War Hero, First American Dog to Enter German Territory.”

St Augustine defined virtue as “the due ordering of the affections,”10 in which everything is loved ordinately, i.e. it is accorded just that love proper to it. Such an understanding of virtue is predicated on an objective hierarchy of goodness and beauty, and on absolute truth, otherwise it would be impossible to love ordinately, as there would be no objective way of ordering things or one’s love for them. The concept of value, however, permits the individual to love inordinately, as that which is valued is based, ultimately, on a subjective personal evaluation and the qualities which either the individual or wider society values may be devalued as opinions change. I would argue that this inordinate love has contributed to the devaluation of courage, particularly in contemporary popular culture, in which an avant-garde artist producing obscene works might be described as ‘daring’, or a person patiently enduring a terminal illness under excellent palliative medical care is described as ‘brave’, or, as happened in Essex recently, a man surreptitiously photographing a thief with his mobile phone, is hailed as a ‘hero’. Perhaps our immediate forebears would have avoided this devaluation of virtue if they had
heeded Sir Thomas Browne rather than Nietzsche. Sir Thomas, writing in the late 1600s, after the Civil Wars had had a similarly disruptive effect on British society as that of the First World War, advised:

“Live by old Ethicks and the classical Rules of Honesty. Put no new names or notions upon Authentic Virtues and Vices. Think not that Morality is Ambulatory; that Vices in one age are not Vices in another; or that Virtues, which are under the everlasting Seal of right Reason, may be Stamped by Opinion. And therefore, though vicious times invert the opinion of things, and set up new Ethicks against Virtue, yet hold thou unto old Morality; and rather than follow a multitude to do evil, stand like Pompey’s Pillar conspicuous by thyself, and single in Integrity. And since the worst of times afford imitable Examples of Virtue; since no Deluge of Vice is like to be so general, but more than eight will escape; Eye well those Heroes who have held their Heads above Water, who have touched Pitch, and not been defiled, and in the common Contagion have remained uncorrupted.” (Christian Morals, Part I, §12)

Courage in war

The traditional Western conception of courage as being dependent on prudence and justice and a “due ordering of the affections”, raises questions about the contemporary understanding of military courage as, by implication, an act of bravery committed in the context of an unjust war, for example, cannot be truly courageous. Part of our difficulty in understanding this arises because the contemporary understanding of Just War has been completely divorced from any understanding of human virtue. I think, however, that a contemporary attempt to justify the courage of, say, a German soldier fighting for the Kaiser or the Nazis, would be based on the distinction in Just War Doctrine between the *ius ad bellum* criteria of deciding when it is just to go to war, and the *ius in bello* criteria, which guide the just conduct of military operations during a war. Michael Walzer (2006:21), for example, argues that it is possible for an individual to fight justly in an unjust war, and that all soldiers, irrespective of the justice or otherwise of their cause, have “moral equality” if they conduct themselves within the *ius in bello* criteria (Walzer 2006:34f.). Jeff McMahan (2009:6) argues against this point of view, by claiming that soldiers “cannot satisfy the constraints of jus in bello, even in principle, when those constraints are properly understood… it is morally wrong to fight in a war that is unjust because it lacks a just cause.” McMahan’s argument, while superficially similar to the traditional Western view, appears to be based on the *ius ad bellum* criteria of whether or not there is a just recourse for going to war, but in the traditional discussions of Just War as laid out by Augustine and Aquinas, the *ius ad bellum* criteria are presented for the guidance of governors, not ordinary citizens, so the individual combatant has no part in deciding whether or not the government has a just recourse to war, but he does have a moral obligation to obey duly constituted authority. It may be true that it is morally wrong to fight in a war lacking a just cause, but the private citizen – even in a contemporary Western liberal democracy, let alone in Imperial or Nazi Germany - does not have the means of determining whether or not that is the case. The combatant can, however, determine to conduct himself according to the *ius in bello* criteria of the discriminate, proportionate and accountable use of force, as Walzer notes. Nonetheless, it is irrelevant whether the individual German soldier personally fought justly on behalf of the Kaiser or the Nazis, as his conduct still does not provide a satisfactory response to Aquinas’s claim that “the praise of courage is dependent on justice.” Indeed, his personal conduct, loyalty and fighting spirit, merely serve the ignoble end of supporting and sustaining an unjust cause or government. Manfred Rommel, who served in the Luftwaffe during the Second World War, describing the moral dilemma which eventually led his father, Field-Marshal Erwin Rommel, to oppose the Nazis, noted,

“all secondary virtues such as bravery, discipline, loyalty and perseverance only have validity so long as they are used in a good cause. When a positive cause becomes negative, these virtues become questionable. The German army had to experience this bitter truth during Hitler’s régime.” (Rommel 1990:viii)

This is hard for contemporary soldiers to hear, because we feel intuitively that a German soldier in 1914 or 1941 who, say, risked his life to save wounded comrades under enemy fire was no less courageous than a British soldier who carried out the same kind of deed and was awarded a Victoria Cross for his bravery. The dilemma arises not simply from the fact that contemporary moral philosophy focuses on human values and actions, rather than on human virtue and character, as mentioned above, but on a deep instinct that any person who willingly risks his life for others is truly courageous irrespective of nationality, cause or ideology. Perhaps Aristotle’s description of the truly brave man provides the basis of an answer:
The courageous man, therefore, is he who endures or fears the right things for the right reason in the right way at the right time, and who shows boldness in a similar manner. For the courageous man feels and acts as the circumstances merit and as reason dictates."  

The courageous man, then, is the one who, in the face of any specific danger, fears - and loves - ordinarly because, through prudence, he understands the reality of the immediate circumstances, and, through justice, knows how to respond appropriately, despite the danger, to secure the good in those particular circumstances. So, perhaps, an individual soldier can exhibit true courage in an unjust cause, if his risking of his life is virtuous in the immediate context over which he has any influence, regardless of whether or not his government is just or engaged in a just war; something over which he has barely any influence?

Conclusion

According to the traditional Western understanding, the courageous man is the one, who in the face of evil and injustice, is willing to lay down his life to secure justice and the good. In short, only the morally virtuous can be truly courageous, because only they can respond ordinarly in the face of evil, danger and injustice. This conclusion, that only a return to the concept of virtue, as understood by Aristotle and Aquinas, will allow us to recognise and, therefore, to appropriately honour true courage, will not commend itself to contemporary people, because it requires that we take our moral character seriously and actually strive to develop habits of goodness. Interestingly, however, based on his own experiences of the First World War, Lord Moran (1945:170) reached the same conclusion as that of the Graeco-Roman philosophers and medieval theologians:

“I contend that fortitude in war has its roots in morality; that [military] selection is a search for character, and that war itself is but one more test – the supreme and final test if you will – of character. Courage can be judged from danger only if the social significance and meaning of courage is known to us, namely that a man of character in peace becomes a man of courage in war. He cannot be selfish in peace and yet be unselfish in war. Character as Aristotle taught is a habit, the daily choice of right instead of wrong; it is a moral quality which grows to maturity in peace and is not suddenly developed on the outbreak of war. For war, in spite of much that we have heard to the contrary, has no power to transform, it merely exaggerates the good and evil that are in us, till it is plain for all to read; it cannot change, it exposes. Man's fate in battle is worked out before war begins. For his acts in war are dictated not by courage, nor by fear, but by conscience, of which war is the final test.”

Or, as Field-Marshal Montgomery (1963) said:

“There is much talk the world over about freedoms, but there is only one positive freedom - the freedom to choose between good and evil.”

Endnotes

1 "Μὴ δὲν τε ἡ τήτη ῥ Ε μεναι προκτη ρά τε ἐ ῥ γων." (Iliad IX.443)
2 "And smale fowles maken melodye | that slepen al the night with open ye, | so priketh hem nature in hir corages." (Prologue to The Canterbury Tales, II. 9-11)
3 "dulce et decorum est pro patria mori" Horace, Odes III.2.13. Cf. Wilfred Owen’s poem with this line as its title.
4 "fortitudo sine iustitia iniquitatis materia est." (De Officiis I.35.176)
5 "homo non exponit personam suam mortis periculis nisi propter iustitiam conservandam. Et ideo laus fortitudinis dependet quodammodo ex iustitia." (Summa Theologiae II.II.123.12 ad 3)
6 "καλοὶ δή ἐ νεκα θ ἀνδρείας ὑς υπερμεί ας καὶ πράττει τα κατὰ τὴν ἀνδρείαν." (Ethica Nicomachea III.vii.6 1115b22)
7 "Sed ea animi elatio, quae cernitur in periculis et laboribus, si iustitia vacat pugnatque non pro salute communi, sed pro suis commodis, in vitio est; non modo enim id virtutis non est, sed est potius immanitatis omnem humanitatem repellentis." (De Officiis I.xix)
8 "iustitia est constans et perpetua voluntas ius suum cuique tribuens." (Imperatori iustiniiani Institutiones I.1)
9 "Omnis virtus moralis debet esse prudens." (Quaestio Disputata de Virtutibus in Communi 12 ad 23) – cited in Pieper 1966:5
10 "ordo amoris" (De Civitate Dei XV.22)
11 "ό μὲ ν ω ν ἃ δει καὶ ω ν ἐ νεκα τοις ποιένων καὶ φοβούμενων καὶ ὃς δει καὶ ὃ τε, ὁ μοιως δὲ καὶ θαρρών, ἀνδρείας ος κατ’ ἀξίαν γάρ, καὶ ὃς ἡν ὁ λόγος, πάσχει καὶ πράττει ὃ ἀνδρεία ος." (Ethica Nicomachea III.vii.5 1115b17)
Religion in the public arena: 
A strategic challenge facing the Church of England

Rev Jonathan Chaffey, RAF

INTRODUCTION

The relationship between the sacred and the secular in British life has undoubtedly evolved more quickly over the last two centuries than over most of last two millennia. On 7 November 1805 Nelson’s deputy, Collingwood, posted a letter in The Times, giving thanks to God for the victory of Trafalgar and calling for a day of “general humiliation before Almighty God and thanksgiving…imploring forgiveness of sins and a continuation of God’s mercy”¹. Such national acknowledgement of the divine presence in national life contrasts markedly with the disapproval received by Archbishop Runcie from the Prime Minister in 1982, when he dared to propose, in the National Thanksgiving Service for victory in the Falklands War, that war was a terrible thing and that God’s mercy extended to those on both sides of the conflict².

The church in Britain faces the challenge, on a scale not seen perhaps since the emergence of Christianity on these islands through the Celts and Augustine of Canterbury, of fighting for its place in both the public and private spheres. This dissertation’s aim is to address the strategic challenge facing the Church of England (C of E) in the public arena. It will first identify the unique historical and theological heritage of the C of E as a national church that has given it status and reach at all levels of society. It will then conduct a simple SWOT analysis of its current position and influence, exploring both the tensions yet opportunities of the church and state relationship together with threats that emerge from the changing religious, social and philosophical landscape. After highlighting some of the weaknesses of the church’s institutional legacy and its limited resources, it will propose a simple ends, ways and means strategy for the fulfilment of its mission within the public arena in partnership with the state, while looking for evidence of strategic leadership within the church.

This dissertation’s aim is deliberately ambitious in order to achieve a comprehensive overview of the subject, but it is necessarily limited in its ability to address every aspect of the discussion. There are a number of areas not covered in this work that would merit further attention. One concerns the important place of other denominations and religions in the public space. A second area would be an examination of the communication dimensions of strategy, such as the owning of the church’s narrative and its dissemination to wider society and to all members of the church. A third area to explore would be different understandings of the relationship between power and influence. This paper focuses on the church’s strategic challenge, with further research needed for the challenge facing the state in recognising the critically important position of religion in the public arena. The paper is also written with an assumption that the C of E sees itself as called to join in a mission that principally belongs to God, with the Holy Spirit being the primary counsellor and guide through the strategic challenge.

Despite these caveats, the breadth of material covered will allow analysis of fundamental components concerning religion in the public arena, such as the extraordinary reach of the church at national and local level, the fundamental importance for both parties of the church and state relationship, the challenge for the C of E to hold different loyalties in tension, the dominant threat posed by secularism and the creative ways and means necessary to stimulate growth.

THE UNIQUE INHERITANCE OF THE CHURCH OF ENGLAND

In order to understand the current influence of the C of E in the public arena it is necessary to appreciate its inherited status and influence within English society for it has a theology and ecclesiology framed through historical circumstance. The roots of this lie in the community life of the early Christians, with a relationship to the State first outlined by Christ’s preaching of a Kingdom ‘not of this world’ and his
imperative to “Give to Caesar what is Caesar’s and to God what is God’s”\(^1\). It is not the purpose of this paper to trace in depth the theological development of the church’s civic responsibilities, but one could summarise by saying that the embracing of commitment to the state yet with a greater loyalty to God was developed by St Paul\(^2\), given further shape by Augustine’s ‘City of God’ (in the fourth century at a time when Christians were becoming both makers and exercisers of power), then expressed uniquely in the relationship between the C of E and the state through the peculiarity of the English Reformation.

The single most significant event in the C of E’s evolution was not theological but political; the Act of Supremacy, first passed in 1534 when Henry VIII, whilst remaining primarily Catholic in doctrine, broke from Papal Supremacy. Elizabeth I, in her concern to bring a religious settlement to England, assumed the title of ‘Supreme Governor’ of the national church and sought to impose state control of religious practise through the 1559 Oath of Supremacy. This required all clergy (and, later, MPs and members of universities) to “utterly testify and declare in my conscience that the Queen’s Highness is the only supreme governor of this realm, and of all other her Highness’s dominions and countries, as well in all spiritual or ecclesiastical things or causes, as temporal (sic)”\(^5\). Hooker’s ‘Of the Laws of Ecclesiastical Polity’ (1594), in part a response to the puritan and Lutheran emphasis on the polity (or governance) of church life through the laity\(^6\), further served to defend the status of the monarch as an instrument of God’s authority, setting out the position of Anglicanism as the via media between ‘Romanism’ and ‘Popular Protestantism’ long before the term’s coinage by Newman. Hooker saw church and state as two inseparable aspects of the one commonwealth. The fact that clergy are still required on appointment to make the Oath of Allegiance to the sovereign as well as the Oath of Obedience to church authority\(^7\) is testimony to the enduring nature of this intertwined church-state relationship.

Alongside tightly controlled governance, the relationship of church and state was strengthened through a common language and doctrine, the 1662 Act of Uniformity placing the Book of Common Prayer (BCP) as the sole instrument of public liturgy. Together with the 39 Articles of 1563, the BCP Preface fostered a common identity through promoting a broad understanding of doctrine: “It hath been the wisdom of the Church of England…to keep the mean between two extremes, of too much stiffness in refusing, and of too much easiness in admitting any variation from it”\(^8\). Anglican character was further outlined by Hooker’s suggestion of a process of discernment, perhaps even of temperament, which is distinctively Anglican, built on the theological tripod of scripture, reason and tradition.

One could thus argue that the governance, language and character of the C of E, forged through the politics and theology of the English Reformation, has led the C of E to occupy a unique place as a religious denomination within the life of a nation state. The other Provinces of the Anglican Communion, which evolved from missionary and colonial endeavours, hold allegiance to the See of Canterbury but not to the Crown nor to other Heads of State. It has led the C of E to develop a particular liturgy (including the priority of prayers for the monarch and state authorities), an understanding of its mission as serving both soul and body and a decision-making process that is neither papal nor congregational, but rather ‘episcopally led and synodically governed’. Its true distinctiveness, however, lies in its inherited public status and breadth of influence. No area of life is untouched by the relationship, as illustrated by the dramatic symbol of laying military colours on altars, in church parades or at drumhead services. Although critics might argue that such practice somewhat assumes ‘God is on our side’, it serves more purely as a vivid illustration of the state’s recognition of divine presence within the ambiguities and complexities of the military environment. The use of armed force by the state, the ultimate instrument of hard power, is thereby perceived as subject to scrutiny by a higher authority.

Given the inherited relationship of church and state, this paper will now address the current strengths, opportunities, weaknesses and threats that emerge for the C of E.

**STRENGTHS AND OPPORTUNITIES**

The legacy of the C of E’s evolution has given it a unique reach and influence that is both national and local, as witnessed by the 26 Lords Spiritual in the Upper Chamber of the Houses of Parliament and the C
of E strapline: “A Christian presence in every community”. A 2013 report by the think tank, ResPublica, draws attention to the significant endorsement of this level of engagement by the current Sovereign, in a speech at Lambeth Palace in 2012, in which she expressed the broad role of the C of E as a representative both for faith communities and the disadvantaged in society: “Its role is not to defend Anglicanism to the exclusion of other religions. Instead, the Church has a duty to protect the free practice of all faiths in this country”. She went on to highlight the importance of faith in creating and sustaining communities and as a spur for social action. The Secretary of State for Communities and Local Government (the department with a particular remit for faith and community development), in a letter to The Daily Telegraph, signified governmental endorsement for this privileged role:

Christianity in all its forms has shaped the heritage, morality and public life of Britain; and Christian belief continues to influence our society for the better… from the role of Anglican bishops in scrutinising legislation in the House of Lords, through the moral leadership offered by Christian leaders, to the contribution of thousands of churches and Christian charities to the social fabric of our neighbourhoods with their volunteering and sacrifice.

He concluded that the Christian Churches, with their unique position in British society, have a particularly strong claim to be heard, not least in policy-making.

So what does the state want from the church? One can deduce that the state sees the C of E as exercising a clearly defined role in the public arena: representation for all faith groups whilst promoting the Christian religion, policy engagement with government and moral leadership. In the broader community it is looking for partnership with the church in caring for the disadvantaged, and continued voluntary and sacrificial service by the church at a local level.

Within this intertwined relationship it is important to appreciate what the C of E considers it is offering to the state. There is a theological imperative for church involvement in both national politics and local affairs. Hastings refers to Niebuhr’s polarities, ‘Christ against culture’ and ‘Christ of culture’, a theological deduction from Christ’s prayer in John 17 v 16-18 concerning his followers being ‘in’ the world even though they are not ‘of’ it, to suggest that the Church should be neither subsumed by, nor isolated from, the wider society, but rather be sanctified for service within it. This is reinforced by Christ’s social teaching in the Sermon on the Mount, addressing his followers as ‘salt of the earth’, together with his emphasis on the Kingdom of God as a present reality and not just a future hope. Deeper within the Judaeo-Christian tradition is the concept of ‘shalom’, meaning wholeness or completeness for both individuals and communities, with the injunction given to the people of God to seek the peace and shalom of the city. In other words, the text of the Bible calls the church to affirm signs of the Kingdom of God in the culture and structures of society while also retaining a prophetic voice as a critical friend of state authority. Although the biblical injunction is relevant for all denominations, it adds theological legitimacy to the C of E’s historically privileged position within society. By denying any divide between the secular and the spiritual, the church is theologically compelled to help individuals, communities and government to discover and promote the sacred in the ordinary and the infinite in the finite. Partnership with government is thus an opportunity for the church to fulfil its calling.

Current social engagement by the C of E, expressed locally but of national significance, is an example of theology worked out in practice. It is built on a formidable legacy of care, following notable nineteenth century exponents such as Wilberforce and Shaftesbury, pioneers in anti-slavery and factory reform respectively. Brown highlights a modern day example of influence in the policy arena, in the House of Lords debate on the Welfare Benefits Uprating Bill and the letter signed by 43 Bishops, drawing attention to perceived injustices brought about by governmental changes to the welfare system. The church’s engagement has been further demonstrated in recent reform of banking standards and, over the longer-term, in education, with approximately 4500 C of E schools, one million pupils and 22500 governors. The reach and influence of the C of E also extends to individuals and communities to whom the state may have limited access, both in crisis and steady state. During the Tottenham riots of 2011 the church was able to act as intermediaries, treated differently from the authorities, “because the people know that we
live in the midst of them, share their experiences and care about them as individuals. More generally, the ministry of ‘Street Pastors’, working at night with vulnerable people in many major cities, complements the work of other government agencies, such as health authorities and the police, and is a sign of the holistic care offered by a church that offers both practical support and existential hope and meaning. Nor is the involvement limited to those experiencing issues of debt, unemployment and family breakdown in inner-urban areas, but rather extends to the more hidden problems of isolation and immigration in suburban and rural areas, for every piece of England is covered by the parish system. There are also over 1000 full-time C of E chaplains (and many others part-time and volunteer) embedded within industry, the Armed Forces, prisons, schools and hospitals. At a time of austerity and the questionable sustainability of public funding for welfare support, the role of the Church in the ‘Big Society’ cannot be understated.

Interim summary

A complex interweaving of church and state has brought considerable strengths and opportunities to the C of E, not least of which is its unsurpassed access to all areas of national and local life. Embedded within parliamentary structures, with the ability to influence policy and law, it also reaches the most isolated and vulnerable of individuals and communities. With a vast volunteer force, together with an holistic and inclusive approach to care, it offers a considerable service to the nation, comprehensive engagement in the public and voluntary sectors and a bridge-building capacity between the powerful and the vulnerable. It is, nonetheless, a position that carries an inherent challenge, notably the tension of holding two loyalties with integrity. Whilst this can serve as an opportunity for creative partnership, the next section of this paper will address the threats and weaknesses that face the relationship.

THREATS AND WEAKNESSES

A number of challenges face the C of E, generated by the embedded nature of its union with the state but also from sociological and philosophical forces external to that relationship. These threats have exposed potential weaknesses within the church’s internal life and external mission.

Creative tension or divided loyalty?

Creative close relationships invariably require the expression and resolution of conflict. Williams draws out the tension that arises from Augustine’s thesis that the identity and integrity of the Church are not contingent on the success of any particular political order. Graded levels of loyalty mean that the degree of acceptance of an individual to state authority is subject to a higher loyalty to God on issues concerning fundamental moral choices, which is worked out both personally (one might term this ‘protestant’) and through the church (‘catholic’). This opens the door for what Grint calls ‘constructive dissent’. Grint sees balanced and creative decisions within organisations as more likely within an environment in which leaders and followers are encouraged to engage in robust debate. One could equally surmise that the promotion of economic prosperity for the common good and political leadership as service is more likely in a society that promotes a relevant and integrated religious and moral voice in the public arena. Constructive dissent, carrying more creative possibilities than constructive assent, and certainly more collaborative than destructive assent and dissent, requires mutual respect and trust. In one sense the C of E exercises strength in this area, being well-placed through its historical investment in education, its position within the parliamentary process and its customary role as a voice for all religions within the UK, in challenging politicians to face profound questions of meaning when drawing up public policy. Yet it is possible to overstate this strength. Although the C of E’s public opposition in 2013 to the Prime Minister’s Marriage (Same-Sex Couples) Bill sharpened the moral discussion on the purpose of marriage, whilst ensuring opt out clauses for the churches, it failed to affect the outcome of the public debate. One could conclude that the Church, even when it ‘speaks truth to power’, has only limited influence if political will is hard set in a different direction.
Critical solidarity, with the juxtaposition of potentially competing loyalties, can risk destabilising a relationship, as demonstrated by ecclesiastical questioning of the morality of armed intervention in Iraq in 2003. It also cuts both ways, opening the door for criticism of the church by the state, as witnessed in the political storm following Archbishop Runcie’s sober and forgiving tone in his sermon at the 1982 Falklands Memorial Service. This constraining relationship is further demonstrated legislatively, as when Parliament rejected the Church’s proposal for a new Prayer Book in 1928. The state’s criticism of the church’s involvement in public affairs may thus be considered an external threat, to the extent that the former has the power to withdraw the privileged access and influence of the latter. The state could, for instance, choose to remove the Lords Spiritual or faith-based education. At the same time, this could be seen as strength, if the state has sufficient moral standing to be a credible, critical friend towards the church. When it does not, then the fault-line in the relationship is revealed, as with St Peter and St John before the Sanhedrin: “Judge for yourselves whether it is right in God’s sight to obey you rather than God. For we cannot help speaking about what we have seen and heard.” In more recent times, the German pastor, Bonhoeffer (amongst others), when confronted with the evils of the Nazi state, recognised his civic responsibility to disobey, leading to his involvement in the 1944 bomb plot against Hitler and his subsequent execution.

Secular critics (such as the National Secular Society) of the C of E’s established position point to its privileged access to power whereas criticism from within the church might argue against undue influence of the state in church affairs. Both arguments generate an internal threat within the church, which might be tempted to compromise its core mission and values in order to accommodate its public position. Comfort can certainly breed complacency and lead to an unhealthy focus on internal debates that have little resonance with an external audience. Research conducted by the Church Growth Research Programme, for instance, suggests that styles of service and gender of leadership have little impact on church growth, yet these subjects have taken up disproportionate energy within church debates for decades.

One could thus argue that this inherent threat to the integrity of church life is a fundamental weakness, with the church’s mission being better served through disestablishment, a form of ecclesiological ‘conscious decoupling’. Nonetheless, as long as the integrity of the church is maintained, then the theological rationale for church–state engagement, together with the opportunities for influence that establishment brings, suggest that the church’s ministry and the state’s moral wellbeing would be disadvantaged through separation. The challenge for the church of holding different loyalties will always be present; if it were not, then either the church would have been subsumed within the values of the earthly city or become separatist by concentrating on a future (not present) heavenly city. The integrity can be held. Archbishop Welby has put it succinctly: “It’s not about politics; it’s about love.”

The greater strategic challenge facing the C of E comes not from its inherited position, but from two fast changing contextual factors, namely the religious landscape within British society and the growth of postmodernism and secularism.

The challenge of a changing social and religious landscape

It could be argued that the phrase ‘Christian Britain’, referring to a society founded on commonly agreed Christian beliefs and values, is no longer applicable. Certainly, in terms of numerical adherence to church membership, there has been a considerable drop in church attendance since the Second World War. Brown estimates that church adherence fell by almost half and church attendance by a third, just in the years 1974-2000. The related demise of the Sunday School movement has meant a loss of ‘the chain of memory’, the transmission of Christian culture between generations.

The roots of this decline can be linked to the continued industrial and technological revolution within the UK since the mid-Victorian era, with the consequent enormous changes in employment, social structures, mobility and culture that have taken place, on a scale and at a speed, arguably, that is unprecedented. Social and economic developments within the twentieth century that had a tremendous effect on the
influence of the Church include the widespread disillusionment with religion that ensued from the First World War (the supposed ‘war to end all wars’), the economic depression of the 1920s and early 1930s, post-war austerity, followed by the ‘swinging sixties’ and finally the ‘dot com revolution’. Each of these subjects merit considerable study in their own right, but are highlighted here to demonstrate the nature and pace of change.

The government publication, Social Trends 41 reveals the social results of these changes in the second half of the 20th Century. Three examples highlight the challenge for a parochial system that has relied on stable geographical and demographic trends: greater mobility, due to the ease of travel and to economic necessity; a changed working pattern, with a ‘career for life’ becoming increasingly rare; and a more gender-equal labour market, reflected by the rise in the average age of first child-birth from 24 in 1971 to 29.5 in 2010.

An additional challenge to the national and local influence of the C of E is the increasingly pluralist nature of British society. In the face of the apparent demise of inherited Christian faith, there has been a rise in non-Christian religious affiliation in the UK. The 2011 Census for England and Wales records that there was a decrease in people, between 2001 and 2011, who identified themselves as Christian (from 71.7% to 59.3%); yet in the same period there were increases in the other main religious group categories, with the number of Muslims increasing the most (from 3.0% to 4.8%). Furthermore, HMSO statistics for 2009 reveal that in 2008/09, 32 per cent of adults aged 16 and over in England and Wales who reported being Christian, actively practiced their religion, compared with 80 per cent of those who reported being Muslim.

Although this is undoubtedly a challenge to the pre-eminent position of the C of E in national religious life, and one that forces the church either to compete for allegiance or to find accommodation for different faith voices in the market place, it does not question the role of public religion in itself. The greater threats come from philosophical sources, those of post-modernism and secularism, the first one targeting traditional assumptions of religion, the second its place in public life.

**The challenges of postmodernism and secularism**

Cray suggests that identity in pre-modern times was founded on a shared sense of being rooted in the past. In modernity one could fashion one’s identity through profession, family and a common sense of progress. Postmodernism, however, brings a lack of confidence either in the past or in the future. Instead of knowledge being good and offering certainty and objectivity, it is largely discredited. Postmodernism offers few roots in the past and rejects overarching traditions, such as Judaeo-Christian redemption history. Instead, context and perspective are all-important, with no acceptance given to the idea of an ontological reality, only the reality that we ourselves construct. Identity is less inherited or assumed, but is personally constructed through choosing from a wide selection of ideals and lifestyles. Consumption rather than production is the dominant mode of economic organisation, feeding a culture that defines personal freedom as the ability to consume and which regards choice as its core value – “Tesco ergo sum”. Feelings predominate over facts, subjectivity over objectivity, with no absolutes. The resultant liberation, or perhaps confusion, brings a reliance on shifting communities of temporarily like-minded people. Post-modernism therefore rejects the dominant, progressive features of modernism, such as capitalism and a single world economic order, as being patriarchal. It is no wonder that authority structures and figures, such as political parties, the medical profession and the Church, which represent the pillars of modernity, no longer command unquestioned respect. In suggesting that more people find identity, community, morality and mystery in a restaurant than in a church, Drane exposes the challenge to the models and styles employed by the C of E.

Alongside this philosophy that promotes individual choice, Williams outlines the dangers of ‘programmatic secularism’, by which religion is relegated from the public to the private space and where religious belief is seen as secondary to public loyalty to the state. He contrasts this policy affecting approach with a more benign and inclusive ‘procedural secularism’ of a government that seeks to
oversee a variety of communities of religious convictions. Programmatic secularism, translated into policies in education, health, law and banking regulation, cannot deliver on existential or transcendent questions because it robs society of its religious undergirding. Secularisation could, therefore, be described as the process of social change in which formal religion loses social or political significance. It contributes to a loss of belief, together with a declericalisation and fragmentation of religion, through simply not accepting it as part of public debate.

Some observers would, of course, welcome this development. The then president of the British Humanist Association (Polly Toynbee), in a 2010 article in The Guardian, advocated that the Lords Spiritual should no longer have reserved seats in the House of Lords: “The idea that faith offers some missing moral dimension to politics is offensive. All politics is about moral choices. As individuals there are good, wise and clever people of all faiths and none.” Toynbee may have been right in suggesting that religion does not hold a monopoly on wisdom; Christianity would argue the same, given its belief in the God who breathes his Spirit over the whole creation. A significant consequence of the relegation of religion to the private space, however, is the possibility of moral issues of public importance being analysed by ‘how’ and ‘what’ questions, with decisions being made on short-term, utilitarian or popular grounds, thereby missing the all-important significance of the more considered ‘why’ and ‘what for’ questions. The voice of different religions, and especially the voice of the C of E with its historical importance in the framing of national society, needs to be heard in attempting to answer a number of fundamental questions. From where do we derive our moral understanding? Most importantly, how do we nurture and sustain the resilience (individually and corporately) to apply moral values to twenty-first century challenges? These questions were posed to me by a senior military commander, who recognised the importance of religious conviction in engendering moral integrity in the most pressured of environments for personnel of all ranks in the Armed Forces.

How real is this threat? There are certainly elements of programmatic secularism within UK society. One prominent example is the removal of religious education as a core subject in the English Key Stage 3 curriculum. Citizenship education, on the other hand, is compulsory. Whilst some will see this as an inclusive and non-directive development and thus a measure of progress, one could easily argue the opposite. A society that forgets or denies the religious side of life bears the risk of becoming spiritually poor and morally rootless. Equally, to deny the voice of one religious group in order not to offend another actually disadvantages both and ‘dumbs down’ religion as a whole. This is not to deny the potential dark side of religion, in which extremist dogma and practices have led to exclusivism and ethnic violence, but rather to suggest that when a major ethical or political issue emerges within a society which lacks mutual accountability between democracy (or another form of government) and religion, the lack of a constraining and creative religious voice with accompanying moral understanding severely weakens the prospect of good society. Augustine might argue that these ‘two cities’, of the Kingdom of God and a secular power with many ‘gods’, are ultimately irreconcilable, but he also acknowledges their mutual engagement to be important for the prosperity of society: “Even the Heavenly City in her pilgrimage here on earth makes use of the earthly peace and defends and seeks the compromise between human wills... in so far as may be permitted without detriment to true religion and piety.”

**Interim summary**

A consistent reduction in church attendances due to socio-economic and philosophical factors merits a considerable external challenge to the C of E. It is no surprise that the church has lost ground, as society itself has struggled to adapt to a number of strategic shocks, most notably industrialisation, two world wars and the worldwide web. The church, in common with other institutions, has struggled to adjust its structures, hone its focus and translate its message within a changing public arena. In terms of influencing government policy, to some extent the church is at the mercy of the philosophical ‘direction of travel’ within government; it can at least continue to work hard at every opportunity to show the value of its voice and the impact of its practical action. This external risk is made greater, however, by the internal threat of the established church resting on its laurels. By focussing on maintenance rather than mission an inwardly focussed church is likely to suffer a lack of strategic planning. This is further...
complicated by the inadequacy of inherited resources, namely a reduced number of clergy and a surfeit of large, inflexible and costly buildings.

The figures of church decline speak for themselves. The 2011 Census records a growing sense of irrelevance attached to religion, with 14.1 million (a rise from 14.8% to 25.1% of the population) declaring that they had no faith and 7.2 million choosing not to answer the question. Either the population has in general chosen to see religion as increasingly irrelevant or the C of E, as the lead player in the religious field, has failed sufficiently to reinterpret its message in line with societal change. When met by the inflexibility of a church structure that has largely struggled to adapt its hierarchical and geographically based organisation, and to re-imagine its models and styles in making its message relevant to an increasingly disinterested audience, this challenge becomes a crisis.

The church has weathered strategic shocks before, showing adaptability in its methods. Schama cites the Bishop of Bath and Wells writing to his clergy in 1349, during the Black Death. “Priests cannot be found for love or money…to visit the sick and administer the last sacraments of the church”. Given these desperate straits, the Bishop allowed the laity to confess their sins to another lay person or even “to a woman if no man is available”. Unless today’s strategy evolves with fresh thinking and models of practice, then one must conclude that the numerical decline of the church will continue unchecked. The next session will address this challenge by suggesting an alignment of the C of E’s strategy within a simple framework of ends, ways and means.

MEETING THE STRATEGIC CHALLENGE

**Jesus went throughout Galilee, preaching the good news of the kingdom and healing every disease and sickness among the people.**

Paul then stood up in the meeting of the Areopagus and said, “Men of Athens, I see that in every way you are very religious… I even found an altar with this inscription: to an unknown God. Now what your worship as something unknown I am going to proclaim to you.”

Even a cursory reading of the New Testament reveals a very clear strategy for Jesus and his earliest followers. It was both an approach and an activity. They went into the villages, towns and central councils of the Greek and Roman world to proclaim by their words, community life and miraculous signs, that the Kingdom of God had come in the person of Jesus. This fundamental message has not changed over the years or across cultures, nor has the principle of engaging all walks of life, from oppressed individuals to powerful rulers. The strategic vision is to see individuals and communities transformed by the truth and grace of the Christian message.

**The C of E strategy**

Based on the historical and theological evolved position of the C of E, one can suggest a clear ends, ways and means strategy for the church in the public arena. Within this framework the ends are to proclaim and interpret the Kingdom of God in each generation. The ways are engagement, presence and influence at all levels of society, from national to local. The means consist primarily of its people, but also its inherited position, its financial and material assets and its particular strengths of pastoral care for all (regardless of faith), in partnership with other agencies within the public and private sectors. The strategy should be seen more as an activity and approach than a single plan. The ends are not limited by time or context; the ways and means, however, need to be flexible, taking full account of historical position yet changing circumstances and cultures. This matches a definition of strategy given at an RCDS lecture (non-attributable): “A constant process of cohering our aspirations to external challenge using our resources”.

To what extent is this strategy visible within the C of E? The Archbishop of Canterbury, by virtue of his role and his strategic expertise from an earlier career in the banking world, is well placed to give strategic
direction. On taking office in March 2013 he immediately outlined his three key priorities: the renewal of prayer and the Religious Life (meaning religious communities of shared life, work and prayer), as the foundation of ministry; reconciliation, both within and outside the church; and evangelism and witness, as “the only response there is to what God has done for us through Jesus Christ.” Each of these constitutes an end in itself, but together they are also ways to achieve the greater strategic aim. They are not the isolated aims of one individual but are owned by the wider leadership of church and are thus more likely to be driven forward. They build on the C of E’s 3 Quinquennium Goals of 2010-2015: contributing as the national church to the common good; facilitating the growth of the church; and re-imagining the church’s ministry. These corporate goals fit well into the suggested ends, ways and means framework. Contributing to the common good relates to the presence and influence in the public arena. The latter two priorities concentrate on the means of facilitating this strategy, through re-training of the clergy and laity and the development of a ‘mixed economy’ of church ministry, with traditional and pioneering expressions each given support.

So a strategy is visible on paper. It is also possible to observe an increasing sense of intentionality in the speeches and actions of the senior leadership of the C of E. At the first meeting of Archbishops’ Evangelism Task Group in March 2014, the Archbishop of York demonstrated a singularity of purpose: “Next to worship, witness is the primary and urgent task of the Church. Compared with evangelism everything else is like rearranging furniture when the house is on fire. Making disciples is at the heart of our Christian faith and our Anglican tradition.” This mirrors the findings of the Church Growth Research Programme, which has highlighted a clear mission and purpose, together with intentionality in prioritizing growth, as common ingredients of growing churches in any size or context.

The major challenge: Coordination of ways and means

The major challenge lies in matching ways and means; in other words, to consider what is achievable in the current context with available resources. Recent research by the Resource, Strategy and Development Unit of the C of E helps to identify the resource gap. The church is stretched, partly as a result of an economic downturn in society, made worse for the weakest by accessible pay-day loans, which has placed greater demands on parishes to provide social outreach programmes such as food banks and credit unions. At the same time, the downturn has put pressure on the internal raising of funds within churches, with parishes operating in deficit in the last few years and an increasing number using reserves to pay their diocesan contribution. A further diminishing resource is that of stipendiary clergy, predicted to decline by 15% in the next decade, having reduced on average by 1% in each of the last 30 years. When these trends are examined alongside the growing average age of the churchgoing population (higher than the national average), the lack of church attendance in the most deprived (urban priority) areas, churches that are often too large and in the wrong places and the need for training of church leaders and congregations in being more collaborative and outward-looking, one can see the urgency of the strategic priorities for 2010-15, set by the Archbishops and endorsed by the General Synod.

The resource gap raises the question as to whether the ‘Christian community in every community’ is more aspirational than realistic. Statistical evidence, however, suggests that all is not lost for this fundamental principle of the C of E, with signs that the long decline in church attendance is leveling off. The C of E’s annual statistics for 2011 confirm that it still marries 1000 couples, celebrates 2600 baptisms and conducts 3000 funerals every week of the year. This reveals a strong residual interest in what Niebuhr calls ‘primary religion’, the expression of religious ritual during the profound experiences of life. Although the 2011 Census showed a 12% drop in religious affiliation across the population (to 68%), which secularists might attribute to a growing skepticism or rationalism, research by Theos reveals a deeper current of spiritual awareness, with 77% of adults and 61% of non-religious people believing in a spiritual world. One extraordinary example of this phenomenon is recounted by Padre Sharkey, based on his experience as a chaplain in Helmand Province. While conducting a vigil service for a company in a Forward Operating Base, he was astonished to witness the level of concentration on the service with the sangar guards ceasing fire during the silence after the Last Post, even while under fire themselves,
suggesting truth in the axiom that there are very few atheists in a fox hole! In other words, there is openness to things of the spirit that the church can harness; it just needs to be intentional about its aims and interpret appropriately in different settings.

Examples of this adaptability may be seen in the C of E’s use of two key resources, namely young ordinands and the use of material assets. 2012 saw the highest number of young people (aged under 30) accepted for training for the Church of England ministry since 1992, with 113 being 22% of the total.59 The recruiting of younger ordinands who might relate more readily to a younger audience is an example of raising resources to meet need. A greater development of lay (non-ordained) leadership might also provide specialist skills in the mixed economy of church life. Secondly, a study of 42 cathedrals has revealed a 35% growth in weekly attendance and a doubling in weekday attendance between 2002 and 201260. Weekday growth, in particular, signifies a flexibility of approach, offering accessible worship for contemporary lifestyles; in other words, new ways using old resources creatively. Other examples of churches using their buildings to access the postmodern generation (such as Café Church) reflect the latter’s emphasis on image rather than word, on stories rather than facts and mystery over certainty. Transferred into church services, it reflects Finney’s concern that evangelism should be less about doctrine and more about spirit; less a ‘Damascus road’ (conversion event), more an ‘Emmaus road’ (journey of discovery); with mission viewed more as an approach and less as an event61.

The clear goals set by the C of E to address its strategic goals and evidence of fresh growth suggests that church leadership is engaged with the strategic challenge of coordinating ways and means in changing contexts. This will require both moral courage, for difficult decisions to be faced concerning the sustainability of keeping a footprint in every community and also mission command, with intent outlined centrally but authority being devolved for its execution. One current pressure point concerns the allocation of communal funds by the Church Commissioners to poorer dioceses, for distribution to churches with the greatest need62. A Bishop from a northern diocese that is characterised by dispersed communities, with fewer clergy and without strong historical reserves, revealed two concerns63: firstly, that future allocations could be decided through a bidding process to the centre, disadvantaging those dioceses which had fewer resources with which to launch a bid; secondly, that this process was part of a developing philosophy, described as ‘market-based’, which would reward signs of growth rather than supporting the weak. Strategic resolution of these concerns would delegate autonomy to regional dioceses in order to take full account of local culture and need, whilst maintaining ‘buy in’ and accountability at each level of the organisation. If the dioceses continued to work at the operational level, co-ordinating ways and means, then it could still be argued that the resourcing of recognised growth constituted good operational planning. Historically represented by the central Minster, which sent out priests to the villages, this model could be seen as a ‘hub and spoke’ model of church organisation, whereby the strong hub is further resourced in order to strengthen the outlying spokes.

Strategic partnership, not just within the church but between church and government, is an example of the whole being greater than the sum of its parts. It is also essential given the scale of the challenge, not least for government in trying to meet the growing welfare burden, created by a requirement by government to reduce government debt, leading to cuts in real terms across the public sector, together with an ageing population. This is not unique to the UK, the German Chancellor making the stark claim in an interview with the Financial Times (16 December 12), that Europe accounts for just over 7% of the world’s population, 25% of its economy, and 50% of global social welfare spending64. The C of E is not able to offer an alternative to the welfare state and has no wish to see a properly comprehensive response to basic material needs, which is the proper responsibility of government, replaced by a patchy and unpredictable range of charitable responses65. It is, however, ideally positioned to work with Westminster in fostering a moral and just policy of welfare reform, and to work in partnership with government agencies at the community level. This partnership is demonstrated in the Near Neighbours programme, funded jointly by government and the Church Urban Fund, fostering projects to build community in ethnically and religiously diverse inner cities66.

So does the strategy require evolution or revolution? It is clear that the strengths and opportunities of
the C of E at national and local level give wide scope for the exercise of its core ministry. If its ways of working are carefully harnessed, with full engagement with the instruments of state, and if mission command is embraced with a willingness to evolve radical and traditional models of ministry that reflect the culture and needs of local communities, then it has a strategy that will continue to be relevant in the public arena.

CONCLUSION

The C of E has a privileged position within the public arena, its historical and theological evolution giving it a status and reach that is both national and local. The strengths of its inherited relationship to the instruments of state, and at grass-roots level allow it to exercise political and communal influence based on moral and spiritual imperatives. This relationship carries the inherent risk of losing the distinctive integrity and identity of the church’s mission, as the church grapples with its human and divine loyalties. While some would claim that these are incompatible, being in fundamental opposition, this paper has argued that the advantages of critical solidarity outweigh the potential for existential compromise. Indeed, it has suggested that the church, as a representative of the Christian faith and as the primary religious voice in society, has a duty to bring its influence to bear in order to maintain and strengthen the moral and spiritual foundations of society. Through its denial of a divide between the sacred and the ordinary, its inclusive understanding of mission and its reach into all corners of the land, it offers a unique service to the nation.

The strategic challenge to the C of E comes not primarily from a changing social and religious landscape, which has witnessed a long period of decline for institutional religion, although the church may have been slow to adapt to these changes and to translate its message in new ways and contexts. Nor does the challenge come from pluralism, for this does not threaten the place of religion. This paper has described the principal threat as that of programmatic secularism, which seeks to relegate religion to the private sphere. Secularists may use statistical evidence of decline and the institutional inertia of the church as supporting arguments, but the real danger is its presumption that a rational, secular state has no need of communal, spiritual values. This paper has argued that a government that failed to acknowledge fully the church’s voice and message would be unable to provide answers (and may not even ask the question) concerning the fundamental questions of life, and would be unwittingly entertaining an existential threat to public society.

The C of E has clearly appreciated the challenge. Although it carries the weight of some outdated material and diminished financial assets, whilst needing to retrain its primary resource of people for the changing context, it does have strategic direction. For instance, the overall vision of proclaiming and interpreting the Kingdom of God has not needed to change. Secondly, its strapline of ‘a Christian presence in every community’, a key component of the C of E identity, appears to be sustainable, even if it needs to take shape in imaginative new forms. Thirdly, senior leadership, expressed in the key individuals (two archbishops) and the councils of the church, has set out clear priorities in order to achieve the aim. The priorities of prayer and evangelism, for instance, are fundamental to the Christian, resourcing and transmitting the faith. The aims of contributing as the national church to the common good and reconciliation appreciate the distinctive role and potential of the C of E in all levels of society, especially in partnership with state authorities. Re-imagining the church’s ministry shows awareness of the need for realignment of resources.

Bishop Croft, in his commentary on ‘Challenges for the Quinquennium’, notes that the transmission of faith in the face of secularizing culture is the single most serious challenge the church will face in the next generation. In outlining the fundamental need of the church to be open to the grace and guidance of God, whilst listening to the insights of others, he suggests the style of leadership required: “We need to be realistic about the challenges. We need to practise and live hope as a virtue of leadership…we need to prioritise thinking and reflection around this issue. We should beware of simplistic rhetoric and easy solutions.” With leadership that appreciates the strategic challenge, that understands strategy more as an approach than a prescribed series of projects and that is prepared to engage the ‘heavenly city’ with
the ‘earthly city’, this paper concludes that there is indeed hope for the Church of England.

Endnotes

1 Souvenir Issue, The Times, 21 October 05.
4 See, for instance, his teaching on citizenship (Romans 13 v 1-5) and prayer (1 Timothy 2 v 1-6).
14 Matthew 5 v 13, op cit.
15 Amongst several texts, Matthew 4 v 17: “Repent for the Kingdom of God is near.”
16 Jeremiah 29 v 7, op cit.
20 An officer from Police Scotland commented to the author of this paper, whilst on patrol in January 2014, that the police would struggle to manage the nightlife in Glasgow without the Street Pastors scheme. The police are asking for the scheme to grow.
24 Acts 4 v 19, op cit.
34 Rowan Williams, op cit, p 1.
36 Graham Cray, op cit, p 9.
37 An example might be India.
40 Education is mentioned here, but other public sector areas could be cited; for example, the perennial struggle of hospital chaplaincy to justify its existence.
44 Matthew 24 v 23, op cit.
45 Acts 17 v 22-23, op cit.