THE OXFORD HANDBOOK OF

CYBER SECURITY
In Memoriam David Upton
# Contents

*Foreword, Sir David Omand GCB*  
*List of Figures*  
*List of Tables*  
*List of Contributors*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td><strong>PART I. CYBER SPACE: WHAT IT IS AND WHY IT MATTERS</strong></td>
<td></td>
</tr>
<tr>
<td>1. The Origins of Cyberspace</td>
<td>7</td>
</tr>
<tr>
<td>David J. Pym</td>
<td></td>
</tr>
<tr>
<td>2. Opportunity, Threat, and Dependency in the Social Infosphere</td>
<td>32</td>
</tr>
<tr>
<td>Greg Austin</td>
<td></td>
</tr>
<tr>
<td>3. A Political History of Cyberspace</td>
<td>49</td>
</tr>
<tr>
<td>Madeline Carr</td>
<td></td>
</tr>
<tr>
<td>4. Cyber Power in International Relations</td>
<td>66</td>
</tr>
<tr>
<td>Tim Stevens and Camino Kavanagh</td>
<td></td>
</tr>
<tr>
<td>5. Ethical Standards and ‘Communication’ Technologies</td>
<td>82</td>
</tr>
<tr>
<td>Onora O’Neill</td>
<td></td>
</tr>
<tr>
<td><strong>PART II. SECURITY IN CYBER SPACE: CYBER CRIME</strong></td>
<td></td>
</tr>
<tr>
<td>6. Cybercrime: Thieves, Swindlers, Bandits, and Privateers in Cyberspace</td>
<td>89</td>
</tr>
<tr>
<td>Roderic Broadhurst</td>
<td></td>
</tr>
<tr>
<td>7. Making Sense of Cybersecurity in Emerging Technology Areas</td>
<td>109</td>
</tr>
<tr>
<td>Claire Vishik, Michael Huth, Lawrence John, and Marcello Balduccini</td>
<td></td>
</tr>
</tbody>
</table>
19. Cybersecurity Information Sharing: Voluntary Beginnings and a Mandatory Future  
   STUART MURDOCH

20. Data Privacy and Security Law  
   FRED CATE AND RACHEL DOCKERY

21. The ‘Insider Threat’ and the ‘Insider Advocate’  
   MIKE STEINMETZ

**PART VI. PERSONAL CYBERSECURITY**

22. Personal Protection: ‘Cyber Hygiene’  
   DAVE CLEMENTE

23. Online Child Safety  
   JOHN CARR

24. Educating for Cyber (Security)  
   ROGER BRADBURY

   JONATHON PENNEY

**PART VII. NATIONAL CYBERSECURITY**

26. Securing the Critical National Infrastructure  
   DAVID MUSSINGTON

27. The Role of Defence in National Cybersecurity  
   MIKA KERTTUNEN

28. Cybersecurity Capacity Building  
   LARA PACE AND PAUL CORNISH

**PART VIII. GLOBAL TRADE AND CYBERSECURITY**

29. Cybersecurity, Multilateral Export Control, and Standard Setting Arrangements  
   ELAINE KORZAK

30. Cybersecurity, Global Commerce and International Organizations  
   DAVID FIDLER
   Franz-Stefan Gady and Greg Austin  

PART IX. INTERNATIONAL CYBERSECURITY

32. Semi-Formal Diplomacy: Track 1.5 and Track 2  
   Nigel Inkster  

33. States, Proxies, and (Remote) Offensive Cyber Operations  
   Tim Maurer  

34. Getting Beyond Norms: When Violating the Agreement Becomes Customary Practice  
   Melissa Hathaway  

35. International Law for Cyberspace: Competition and Conflict  
   Thomas Wingfield and Harry Wingo  

PART X. PERSPECTIVES ON CYBERSECURITY

36. Community of Common Future in Cyberspace: The Proposal and Practice of China  
   Tang Lan  

37. Look West or Look East? India at the Crossroads of Cyberspace  
   Arun Mohan Sukumar  

38. Cybersecurity in Israel: Strategy, Organization, and Future Challenges  
   Lior Tabansky  

39. The Evolving Concept of the Japanese Security Strategy  
   Yoko Nitta  

40. Contextualizing Malaysia’s Cybersecurity Agenda  
   Elina Noor  

41. The Russian Federation’s Approach to Cybersecurity  
   Anton Shingarev and Anastasiya Kazakova
## PART XI. FUTURE CHALLENGES

42. Rethinking the Governance of Technology in the Digital Age  
   Joëlle Webb  
   687

43. Maturing Autonomous Cyber Weapons Systems: Implications for  
    International Cybersecurity  
    Caitríona Heinl  
    702

44. The Future Human and Behavioural Challenges of Cybersecurity  
    Debi Ashenden  
    723

45. The Future of Democratic Civil Societies in a Post-Western  
    Cybered Era  
    Chris Demchak  
    735

46. Future Normative Challenges  
    Eneken Tikk  
    751

47. ‘Cybersecurity’ and ‘Development’: Contested Futures  
    Tim Unwin  
    769

48. Project Solarium 1953 and the Cyberspace Solarium  
    Commission 2019  
    Mike Steinmetz  
    785

Conclusion  
   Paul Cornish  
   798

Index  
   807
Sir David Omand GCB

The publication of this *Oxford Handbook* could not be more timely. Security problems are multiplying in cyberspace at the same time as recognition is dawning of our vulnerability to cyber exploits and cyberattacks, and the large costs that they can impose on governments and businesses. Not only is there an urgent need to apply in practice the analysis of cybersecurity contained in the chapters of this *Handbook*, there is an overdue need to invest more in research into the gaps in our knowledge and understanding, apparent too from the chapters here, to enable us to be prepared for the inevitable development of the threat landscape and the malicious skills of those who populate it.

Recognition of the dependence of modern societies on the digital world of the Internet and the Web it carries, and of the Big Data it generates, has come not before time. Our reliance has crept up on us little by little. Our understanding of the consequent risks we are running has not kept pace, and has become divorced from our actual risky behaviours online. The private sector technology companies were the first to recognize the huge commercial advantages of using the open Internet to communicate directly to customers, building global businesses by being able to ignore the artificiality of national borders. Other businesses, including the financial sector, quickly followed to gather marketing information, communicate with customers and the supply chain, and dramatically cut down transaction costs through disintermediation. Yes, there were losses due to traditional frauds being able to be conducted at scale and new forms of digital crime but at first these would have seemed bearable in comparison with the benefits. As losses have mounted, with the easy gains already taken, that complacency is being shaken.

As mobile devices proliferated at ever decreasing real cost, the everyday life of the citizen was not far behind in being transformed by the connectivity of the Internet and the Web with access to family and friends on social media coming to dominate everyday interactions. Upcoming generations can scarcely imagine life without the connectivity of this portable digital environment. Again, the reported online bullying, cyber stalking and trolling at first mostly happened to others; now the prevalence of such personally damaging and destructive behaviour online has grown to disturbing proportions. Russia has demonstrated how digital subversive campaigns can be mounted against the United States and European nations by interfering in elections, exacerbating tensions in society, and undermining confidence in Western democracy.

The development of the Dark Net, exploiting the natural anonymity baked into the founding protocols of the Internet, has led to a proliferation of criminal websites offering
for easy sale every type of illegal recreational drug, guns and other weapons, stolen personal
details, and cyber exploits to enable further criminal fraud and extortion at the expense
of the public. The grooming of youngsters for sexual exploitation on social media is now a
major preoccupation of the police, along with trying to inhibit the sharing of child pornog-
raphy and online streaming of extreme child abuse, often from the other side of the world,
exploiting the borderless characteristics of the Internet. Terrorist groups have used a range
of social media to stream violent propaganda at potential recruits, in ways that make it hard
for the Internet companies to detect and remove it.

In addition, most governments have added to national Internet dependency, taking ad-
vantage of the speed and scale of data technology to cut out back offices and provide citi-
zens with direct access to official services ranging from registering to vote and licensing
motorcars to lodging tax returns. The cost reductions, and gains in convenience to have the
service online, were just too attractive to be delayed in order to allow the massive investments
in designing in security, staff training and developing the technical assistance that would
have been needed to give a high assurance of data security. Just in time, as the Internet of
Things begins to exhibit its potential, we have an overdue greater recognition on the part of
elected governments that there are downsides for the citizen as well as upsides to the digital
revolution.

It was entirely predictable that the wave of digital communications and data storage tech-
nology that developed in the 1990s and 2000s, and is currently cresting and breaking over us,
would bring problems to those used to swimming in shallower waters. Many companies and
individuals alike are now finding themselves out of their technical depth. For many small
and medium-sized companies, the cost of paying for the cyber equivalents of lifeguards and
rescue helicopters on standby in case of trouble is proving prohibitive.

In a historical perspective, those with criminal intent have always been quick to ex-
plot technology for their own selfish benefit, whether using that revolutionary invention,
the motor car, to enable fast getaways from robberies or when commercial aviation came
along to escape jurisdiction, not to mention the use of plastic explosives to blow safes. Today,
there is no need for the criminal to flee across borders because attacks can be mounted over
the Internet from anywhere on the globe. Again, in the past, criminals could usually only
plan and mount one attack at a time; today, the same attempted fraud can be directed at
thousands of customers. And the really big heists do not require the swag to be carted away
because the digital proceeds can be transferred at the touch of a key through a string of
money-laundering accounts.

The ambition of cybercriminals (and their state sponsors in some cases) continues to
amaze, both in its audacity and its global reach. Such was the attempt (thankfully only par-
tially successful) to hack the Swift interbank payment system to defraud a number of central
banks of some $951 million. That attack has been ascribed by cyber investigators to North
Korea (the sum involved would represent one month of Republic’s gross domestic product
and would no doubt have been a welcome addition of foreign exchange funds for intelli-
gence and subversive activity overseas). That attack appears to have had insider help, a re-
minder that cybersecurity is as much about people, and managing them well and with care,
as it is about technology. The case also reminds us that the boundary between criminal and
state activity is not clear-cut. Russian attackers inserted the NotPetya worm into tax prep-
eration software aimed at a Ukrainian target but it escaped into the wild, did over $1 billion’s
worth of damage to global companies, and almost destroyed the world’s largest shipping
container company, Maersk. Digital interference in another country’s internal affairs is now commonplace. The pre–election hack into the US Democrat National Committee, and the subsequent release of stolen emails, intended to sway the 2016 US presidential election away from Hillary Clinton and towards Donald Trump, was a covert action firmly attributed to the Russian authorities and an example of the modern weaponization of digital information. And that attack should also serve to emphasize the value of being able to harness digital and other intelligence sources to assist with the attribution of exploits and attacks. The 2016 attack on the French TV5 Monde television channels, now attributed to Russia, was, for example, conducted with malware compiled in such a way as to try (unsuccessfully in the end) to divert blame on to the so-called ‘Cyber Caliphate’ linked to the Islamic State.

A profession that has always fastened early onto new technological advances is that of secret intelligence. UK naval intelligence was quick to exploit the early development of radio transmissions during the First World War to listen in to the enemy and to geo-locate warships and military units. During the Second World War, it was the advent of proto-computers for deciphering, and the application of mathematics to cryptanalysis, that revolutionized signals intelligence. During the Cold War, it was the development of microelectronics that led to the huge investment in spy satellites. Now we have the global Internet and the World Wide Web, and all the opportunities they offer for bulk access to the communications, and patterns of communication, of the targets of intelligence activity, as well as access through computer network exploitation to the stored data of every kind of organization. The first documented massive computer raid on classified information, Moonlit Maze, was a sophisticated Russian attack on the databases of the US military and research and development sector dating back to the early 1990s and lasting for several years before detection. A hard lesson of the need for cybersecurity protection for national defence networks. Nevertheless, subsequent Chinese so-called ‘advanced persistent attacks’ (APTs) on US and European industry and commerce exposed significant cybersecurity weaknesses in the private sector but not before huge losses of intellectual property had occurred. Finally, the US and Chinese presidents reached an agreement that neither side would conduct cyber espionage on the other for the commercial gain of their national companies, with a comparable agreement reached between the UK and China — rare examples of international cyber norms being agreed.

The increasingly urgent demands today are for intelligence on terrorists, proliferators, and criminals, and these have led to an upsurge in intelligence requirements being authorized for information about people as individuals, rather than the more traditional objects of intelligence attention to support defence and foreign policy. The demand is for the identities, associations, movements, locations, and financing of those who directly mean us harm, and where better to look for such information than the packet-switched networks of the global Internet, in social media use, or in data at rest in bulk personal databases. A cybersecurity dilemma arises at once between the need for the authorities to access such information for public safety and to uphold the law, and the calls for the strongest possible protection against the hacking of mobile phones, tablets, and other devices, and for strong end-to-end encryption to protect the privacy of the individual user and promote confidence in the integrity of the Internet.

The dynamic demand and supply interaction between the need for timely information about people of intelligence and law enforcement interest, and the intelligence opportunities offered by digital technology, often using the very same methods that commerce itself uses to derive marketing information on its customers, has created a strong set of surveillance tools.
The very power of these capabilities, and their potential for misuse in the wrong hands, has led to a significant strengthening in many Western democracies of the constitutional, judicial, and Parliamentary oversight protections for the privacy of individuals and for the maintenance of free speech (in the UK, principally through the Investigatory Powers Act 2016). In totalitarian and less democratic states, the abuse of digital surveillance against dissidents and political opponents is already apparent.

In the long run, a major danger facing the democracies is loss of confidence in the Internet as a secure medium for doing business, communicating with government, and conducting a fulfilling private life. That observation directly translates into the priority now being given by many governments to cybersecurity—for example, in the UK, through the 2016 creation of the National Cyber Security Centre. No one can promise the certainty of future trust in the Internet, and certainly not the elimination of the many cyber risks that exist today. And new risks are certain to emerge—for example, as terrorist groups and malign individuals come to recognize the harm they can inflict on advanced Internet-dependent societies with relatively simple cyberattack tools, many of which are already available to be bought on the Dark Web (for untraceable Bitcoin payment of course).

Most nations are now searching for cybersecurity strategies that will better enable them to manage all these risks. Most strategies contain the elements summed up in the recent euphonious UK cybersecurity strategy of *Defend* against cyber threats, *Deter* all forms of aggression in cyberspace, and *Develop* an innovative, growing cybersecurity industry, underpinned by scientific research and development, with a self-sustaining pipeline of talent providing the skills to meet public and private sector needs.

How such laudable goals are to be met is uncharted territory. Some nations, such as France and Germany, have chosen deliberately to present their centres of cyber expertise as organizationally separate from their intelligence communities, presumably in the hope of increasing public credibility in the security advice being given and to minimize the risk that the search for zero-day exploits to gather digital intelligence will take precedence over securing commonly used software. The alternative view, being taken by the United States and the United Kingdom among others, is that, while good cyber hygiene by companies and the public can secure some 80% of the required security, protecting against the really damaging exploits and attacks requires the information that only digital intelligence can gather, including that to support offensive cyber operations. That logic points to an organizational model that allows national deep cyber expertise to be exploited for both intelligence and cybersecurity purposes.

The policy implications of cybersecurity issues do not naturally fall to any single government department. Home affairs and interior ministries, justice ministries, defence and foreign affairs, finance, trade and business departments all have important interests at stake. Nor is adequate cybersecurity in one country scarcely conceivable in the connected global Internet. Again, this is relatively uncharted territory for the machinery of government and for international organizations.

Governments are naturally uneasy with a situation in which no one is in charge of the global phenomenon that is the Internet. Yet, its very success and rapidity of growth is down to the founding fathers’ insistence on ‘running code and rough consensus’ with the open protocols of the Internet, allowing any compliant network to work seamlessly into the whole and, most recently, for the millions of apps and Internet clouds to spring up for mobile devices, giving users unparalleled computing power at their fingertips. The interests of
cybersecurity as interpreted by some states could quickly put a dampener on the flow of such innovation.

A useful conceptual distinction can be made between cyber exploits, where the intention is to penetrate networks for the purpose of intelligence gathering, and cyberattacks, with the intent to corrupt, deny, or destroy key information (including that in control systems). The coming Internet of Things poses real risks of malicious attacks for financial gain or for malicious damage, as do the myriad control systems that are present at every level of the critical national infrastructure and its key suppliers, almost all of whom are in the private sector. Putting to one side the apocalyptic fantasies of the complete collapse of urban civilization as a result of cyber war, the opportunities for sabotage for the purposes of intimidation, or to place significant obstacles in the way of a state trying to defend itself, add a new dimension to the concept of friction in war. The management of these issues, as many of the chapters in this Handbook discuss, is going to involve a combination of more responsible behaviour by employees and citizens generally, and the introduction of more secure technologies, some of which are still to be developed. What is clear is that no set of security measures that significantly impede people doing their jobs, or enjoying their social lives to the full, is going to succeed because all that will do is encourage even more insecure workarounds.

The Oxford Handbook of Cybersecurity comes, therefore, at the right moment and I hope for all our sakes that it will be widely read, digested, and acted upon in the years to come.
## List of Figures

1.1. The Greek hydraulic semaphore  13  
1.2. A replica of one of Chappe's semaphore towers in Nalbach, Germany  14  
1.3. Spread of the Chappe Telegraph, 1793–1854  15  
1.4. Telegraph Connections (Telegraphen Verbindungen), 1891, Stielers Hand-Atlas, Plate No. 5, 'Weltkarte in Mercators projection'  17  
1.5. Undersea telecommunications cables  18  
1.6. The TCP/IP model and the OSI reference model  20  
1.7. How the Internet implements cyberspace  21  
1.8. Internet users in 2015 as a percentage of a country’s population  21  
1.9. Reasoning about systems and reasoning about system models  26  
7.1. Evolution of cyberattacks  110  
7.2. Cybercriminals and the law  115  
7.3. Anticipated threat themes  117  
7.4. Leveraging Ontologies  119  
7.5. Cybersecurity Metrics and the OODA Loop  123  
9.1. Cyber Incident Lifecycle  144  
9.2. General Model for Vulnerability Mitigation  154  
12.1. Categorization of cyber terrorism  188  
26.1. A Cybersecurity Risk adaptive framework  430  
33.1. Three main types of relationships  550  
46.3. Normative gap analysis  758  
47.1. Conceptualizing the intersection between cybersecurity and development  771  
47.2. Active mobile-broadband subscriptions per 100 inhabitants by level of development. Source: ITU (2018a).  772
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Headline descriptions of classes of dependence for critical infrastructure</td>
<td>43</td>
</tr>
<tr>
<td>6.1</td>
<td>Stolen credit cards – market values circa 2013–2015</td>
<td>96</td>
</tr>
<tr>
<td>27.1</td>
<td>An assessment of defence sector’s contribution to national cybersecurity</td>
<td>455</td>
</tr>
<tr>
<td>31.1</td>
<td>Microsoft’s proposed norms</td>
<td>524</td>
</tr>
<tr>
<td>33.1</td>
<td>Non-state proxies based on organizational structure</td>
<td>546</td>
</tr>
<tr>
<td>43.1</td>
<td>Examples of arguments for possible military or economic advantages and disadvantages of advanced automation/maturing autonomous physical and cyber systems that require validation</td>
<td>709</td>
</tr>
<tr>
<td>43.2</td>
<td>Several technical risks associated with maturing autonomy in cyber systems</td>
<td>712</td>
</tr>
</tbody>
</table>
List of Contributors

Debi Ashenden is Professor of Cyber Security at the University of Portsmouth, UK, and the University of Adelaide, Australia.

Greg Austin is Head of the Programme on Cyber, Space and Future Conflict at the International Institute for Strategic Studies (IISS), and Professor of Cyber Security, Strategy, and Diplomacy at the University of New South Wales.

Marcello Balduccini is Assistant Professor, Department of Decision and System Sciences at Saint Joseph’s University, Philadelphia.

Roger Bradbury is Emeritus Professor of Complex Systems Science at the Australian National University.

José Eduardo Malta de Sá Brandão is Deputy Director of International Studies, Political and Economic Relations at the Institute for Applied Economic Research (IPEA), Brazil.

Roderic Broadhurst is Professor of Criminology at the Australian National University Cybercrime Observatory.

Ben Buchanan is Assistant Teaching Professor at Georgetown University School of Foreign Service.

John Carr OBE is Secretary to the Children’s Charities’ Coalition on Internet Safety, and Visiting Senior Fellow at the London School of Economics and Political Science.

Madeline Carr is Professor of Global Politics and Cybersecurity, and Director of the Research Institute in Socio-technical Cyber Security (RISCS), Department of Computer Science at University College London.

Fred Cate is Vice President for Research, Distinguished Professor, and C. Ben Dutton Professor of Law at Indiana University, and Senior Policy Advisor to the Centre for Information Policy Leadership.

Dave Clemente is Head of Cyber Risk Research, Deloitte, UK.

Alexander Corbeil is Digital Fellow at Concordia University and Sessional Lecturer at Carleton University, Ottawa.

Paul Cornish is Visiting Professor, LSE IDEAS, at the London School of Economics.

Chris Demchak is Professor and Grace Hopper Chair of Cyber Security at the Cyber and Innovation Policy Institute (CIPI), and Senior Cyber Scholar in the Strategic and Operational Research Department (SORD) at the US Naval War College.
Rachel Dockery is Senior Research Fellow in Cybersecurity and Privacy Law at Indiana University Maurer School of Law.

Florian J. Egloff is Senior Researcher at the Center for Security Studies (CSS), ETH Zurich, and Research Associate at the Centre for Technology and Global Affairs (CTGA), University of Oxford.

David Fidler is Senior Fellow for Cybersecurity and Global Health, Council on Foreign Relations, Washington DC.

Franz-Stefan Gady is Research Fellow at the Institute for International Strategic Studies, London.

Sandro Gaycken is Director of the Digital Society Institute at ESMT Berlin and Director of the NATO Science for Peace and Security Programme (Cyberdefence Projects).

Melissa Hathaway is President of Hathaway Global Strategies LLC and former cyber advisor to President George W. Bush and President Barack Obama.

Caitríona Heinl is Executive Director of the Azure Forum for Contemporary Security Strategy, Ireland, and Adjunct Research Fellow at the School of Politics and International Relations (SPIRe), University College Dublin.

Michael Huth is Professor of Computer Science, Department of Computing, at Imperial College London.

Eva Ignatuschtschenko is a former Research Fellow at the Global Cyber Security Capacity Centre, Oxford Martin School, University of Oxford.

Nigel Inkster CMG is Senior Advisor at the International Institute for Strategic Studies, London.

Lawrence John (Major, USAF, retired) is Distinguished Analyst and Industrial Base Systems Engineer at Analytic Services Inc.

Camino Kavanagh is Senior Visiting Fellow at the Department of War Studies, King's College London.

Anastasiya Kazakova is Public Affairs Manager at Kaspersky, Moscow.

Jaclyn Kerr is an Affiliate at the Center for International Security and Cooperation, Stanford University.

Mika Kerttunen is Senior Research Scientist at the Department of Software Science, Tallinn University of Technology.

Elaine Korzak is Visiting Assistant Professor of Cybersecurity at Middlebury Institute of International Studies at Monterey, Palo Alto, California.

Tang Lan is Research Professor and Deputy Director of the Institute of Information and Social Development, China Institutes of Contemporary of International Relations (CICIR), Beijing.
Herbert Lin is Senior Research Scholar (Center for International Security and Cooperation) and Hank J. Holland Fellow (Hoover Institution), Stanford University.

Jon R. Lindsay is an Associate Professor at the Georgia Institute of Technology, School of Cybersecurity and Privacy, and in the Sam Nunn School of International Affairs.

Tim Maurer is Senior Fellow and Co-Director of the Cyber Policy Initiative at the Carnegie Endowment for International Peace, Washington DC.

Stuart Murdoch is the Founder and CEO of Surevine, London.

David Mussington is Professor of the Practice, School of Public Policy, University of Maryland College Park.

Yoko Nitta is Senior Fellow and Russia Study Group Chair at the Japan Society for Security and Crisis Management (JSSC), Tokyo.

Elina Noor is Visiting Fellow at the Institute of Strategic and International Studies, Malaysia.

Onora O’Neill is an Emeritus Honorary Professor of Philosophy at Cambridge, and has been a crossbench member of the House of Lords since 2000.

David Omand gcb is Visiting Professor in the War Studies Department, King’s College London. and a former UK Security and Intelligence Coordinator, Permanent Secretary to the Home Office, Director of GCHQ, and Deputy Under Secretary of State for Policy at the UK Ministry of Defence.

Lara Pace is an independent analyst, has worked with government, academia, and multilaterals in more than 60 jurisdictions.

Jonathon Penney is Research Fellow at Citizen Lab, University of Toronto, and at the Berkman Klein Center for Internet and Society, Harvard University.

David J. Pym is Professor in Department of Computer Science and Department of Philosophy, UCL, and Institute of Philosophy, University of London.

Rafal Rohozinski is Principal, SecDev Group, Ottawa.

Anton Shingarev is Vice President for Corporate Relations, Yandex, and Visiting Scholar at the Higher School of Economics, Moscow.

Mike Steinmetz is Principal, Digital Executive, Limited. Director and General Partner at College Hill Ventures PBC, and Adjunct Professor at Boston College and Woods College. He was formerly State Cybersecurity Officer and Homeland Security Advisor for the State of Rhode Island.

Tim Stevens is Senior Lecturer in Global Security at the Department of War Studies, King’s College London.

Arun Mohan Sukumar is Head of Cyber Initiative, Observer Research Foundation, Mumbai.
Lior Tabansky is Head of Research Development at the Blavatnik Interdisciplinary Cyber Research Center, Tel Aviv University.

Eneken Tikk is affiliate researcher of the Erik Castrén Institute of the University of Helsinki and Executive Producer, Cyber Policy Institute, Jyväskylä, Finland.

Tim Unwin CMG is Emeritus Professor of Geography and Chairholder of the UNESCO Chair in ICT4D at Royal Holloway, University of London.

Nicole van der Meulen is an expert in cybersecurity and former security advisor for the Dutch Banking Association.

Claire Vishik is a Fellow and GMT CTO at the Intel Corporation.

Joëlle Webb is a former Fellow at the Weatherhead Center for International Affairs, Harvard University, 2016–18.

Thomas Wingfield is Deputy Assistant Secretary of Defense for Cyber Policy, United States Department of Defense.

Harry Wingo is a Faculty Member at the College of Information and Cyberspace, National Defense University, Washington, DC.
Cyber security has come to occupy a prominent position in public policy debates (nationally and internationally), in corporate risk analysis, and in the private lives of individual people. Vast amounts of public and corporate effort and investment are devoted to making cyber space more secure from a spectrum of security threats and challenges, including nuisance hacking, insider security breaches, criminality (ranging from low-level theft through to serious, organized activities involving data theft, sexual abuse, and large-scale financial fraud), commercial espionage, political subversion, national security espionage, terrorism and political extremism, and inter-state conflict. Many of these challenges are familiar and need little elaboration. Others are relatively novel, such as the undermining of democratic institutions and their associated legal and electoral processes, the vulnerabilities of health implantables, and the governance of algorithmic decision making. For its part, the public is reminded endlessly of the need for ‘cyber hygiene’ and hears news almost daily of another cyber ‘hack’ or ‘attack’ against national or corporate interests, of foreign interference in domestic politics, and even apocalyptic warnings of imminent ‘cyber war’.

Yet, in spite of these risks and anxieties, the world is becoming ever more deeply dependent on the digital environment. When dependency of any sort is unmanaged and unmitigated, it can soon become the basis of vulnerability. In other words, a vicious spiral is created: as societies, governments, corporations, and individuals become more dependent on the digital environment, so they also become increasingly vulnerable to misuse of that environment, and particularly so when there is an inability or reluctance to close or even merely to manage the vulnerability. Vulnerability is increasingly a function of much of what goes on in the digital environment and, when mitigation is lacking, that vulnerability becomes increasingly brittle, possibly even critical. As well as the threats, hazards, and risks listed earlier, the users of cyber space have therefore become, in a sense, a threat to themselves.

In this ostensibly hazardous, even menacing, environment, it is scarcely surprising that a considerable industry should have developed to provide the means with which to make cyber space more secure, stable, and predictable: threat monitoring and intelligence reporting; hardware and software security; penetration testing; online awareness training; and consultancy of every conceivable sort. The case for the cyber security industry seems compelling enough. On the one hand, the ‘threat picture’ includes adversaries who are highly skilled, aggressive, and fast-moving while, on the other hand, that which they threaten is
becoming ever more valuable to its users. A cynic might argue that cyber insecurity has become a rather compelling and self-serving business case in that, for as long as cyber space is not secure, there is commercial opportunity in managing that insecurity without resolving it. If the cyber security industry is vigorous and expansive, similar things might be said of another sector that has also been developing rapidly. Research and analysis of cyber security as a phenomenon in national and international public policy now receives considerable attention around the world in the media, in think tanks, and in academia. The coverage is wide-ranging and varies in style, quality, and coverage from the highly technical to the policy-analytical to what might, again, be described as self-servingly alarmist.

Cyber security is concerned with the identification, avoidance, management, and mitigation of risk in, or from, cyber space—the risk of harm and damage that might occur as the result of everything from individual carelessness to organized criminality, to industrial and national security espionage, and, at the extreme end of the scale, to disabling attacks against a country’s critical national infrastructure. But this account represents a rather narrow understanding of security and there is much more to cyber space than vulnerability, risk, and threat. As any university lecture on international and national security would point out, the pursuit of security from financial loss, physical damage, etc. should not be seen as an end in itself. Security must also be for something: more than simply the avoidance or elimination of risk, security is also essential for the maximization of benefit. This important point is often explained by analogy. When we lock our front door to protect our house, ourselves, and our property from thieves and predators, we take this action not because we see ourselves as an agent of law enforcement but in order that we can enjoy what we have, live as we choose, and grow as we need. Security, including cyber security, is protective but it is also liberating and enabling. This applies at every level—individually, nationally, and globally—and it could scarcely be more positive and constructive.

It makes little or no sense to approach cyber security from a narrow perspective, categorizing and analysing cyber security according to this or that functional perspective or political level, concerned only with protection from threat, danger, harm, and loss. The security of cyber space is as much technological as it is commercial and strategic; as much international as regional, national, and personal; and as much a matter of hazard and vulnerability as an opportunity for social, economic, and cultural growth. Consistent with this outlook, the Oxford Handbook of Cyber Security takes a comprehensive and rounded approach to the still evolving topic of cyber security. The structure of the Handbook is intended to show that cyber security is far more than a matter of threat, vulnerability, and conflict, serious though these matters are; that it manifests on many (if not all) levels of human interaction; and that an understanding of cyber security requires us to think not just in terms of policy and strategy but also in terms of technology, economy, sociology, criminology, trade, and morality. Accordingly, contributors to the Handbook include experts in cyber security from around the world and from a wide range of perspectives: former government officials, private sector executives, technologists, political scientists, strategists, lawyers, criminologists, ethicists, security consultants, and policy analysts.

With a Foreword by Professor Sir David Omand GCB, former UK Security and Intelligence Coordinator and Director of the UK Government Communications Headquarters (GCHQ), the Handbook comprises 48 chapters organized in 11 parts. Part I—‘Cyber Space: What it is and Why it Matters’ sets out the origins, character, and dynamics of cyber space and explores certain of the moral, intellectual, and practical questions it poses. Parts II to IV
are concerned with cyber space as an environment of vulnerability, threat, and confrontation of three different (though often overlapping) types: 'Cyber Crime' (Part II); 'Extremism and Terrorism' (Part III); and 'State-Sponsored Cyber Attacks' (Part IV). Having examined prominent challenges and threats to security in and from cyber space, the *Handbook* then turns its attention to security-seeking responses and practices, at various levels: ‘Technical and Corporate Cyber Security’ (Part V); ‘Personal Cyber Security’ (Part VI); ‘National Cyber Security’ (Part VII); ‘Global Trade and Cyber Security’ (Part VIII); and ‘International Cyber Security’ (Part IX). The penultimate set of chapters—Part X: ‘Perspectives on Cyber Security’—analyses the approach to cyber security in six countries—the People's Republic of China, India, Israel, Japan, Malaysia, and the Russian Federation. A complete account of national perspectives would, of course, require the inclusion of almost 200 chapters, which would exceed by some pages the capacity of this *Handbook*. It is hoped, nevertheless, that these six chapters will help to offset any bias that might be perceived in the *Handbook* in favour of European and North American views and preferences and, more importantly, will capture something of the wealth of experience of cyber security that has been gained around the world in a variety of environments and circumstances. Finally, in Part XI: ‘Future Challenges’, the *Handbook* invites seven authors to suggest what technological and human challenges might lie ahead.
PART I

CYBER SPACE: WHAT IT IS AND WHY IT MATTERS
CHAPTER 1

THE ORIGINS OF CYBERSPACE

DAVID J. PYM

DEFINING CYBERSPACE

According to The Oxford English Dictionary (OED)¹:

cyberspace |ˈsʌɪbəspiːz|
noun [mass noun]
the notional environment in which communication over computer networks occurs. I stayed in cyberspace for just a few minutes.

According to The Oxford Dictionary of Science Fiction (Prucher 2007):

the entirety of the data stored in, and the communication that takes place within a computer network, conceived of as having the properties of a physical realm;…

My purpose here is to unpack and explain these definitions, which are wholly consistent with each other. In contrast to the approach of Hook and Norman (2002), who achieve an enormously impressive coverage of relevant material, my guiding principle is to explore the sense in which cyberspace is a ‘space’, a concept that is well understood in mathematics and physics, and the understanding of which in those fields is alluded to in the term ‘cyberspace’.

The modern world is more-or-less wholly dependent for its operation on networks of communicating computers. These computers come in all shapes and sizes. They may be small devices embedded in everyday objects such as watches and household appliances, or cars, or personal laptops and workstations, or vast datacentres supporting the infrastructure of cloud computing. What is central to their function is communication—that is, the transmission of data—over local networks, wider corporate or government networks, or the Internet itself.²

In more detail, the Internet consists of a global network of networks that connect computers around the world and use a collection of communications protocols, including TCP/IP (Transmission Control Protocol/Internet Protocol, the basic language that the connected computers use to communicate with one another), and others such as OSPF...
(Open Shortest Path First), BGP (Border Gateway Protocol), and RIP (Routing Information Protocol), which together describe how data should be decomposed into packets (the basic units of data handled by the protocol), addressed, transmitted, routed, and received. Data (which can be treated as a singular, plural, or mass noun) consists of blocks (‘bits’ and ‘bytes’) of binary numbers. Data itself has no inherent meaning, but data is used to represent features of physical or abstract worlds; that is, information.3,4

When thinking of the Internet, it is often tempting to conflate it with the World Wide Web (WWW). While this is quite understandable in many ways—and the distinction is often blurred in common discourse—it is a conceptual mistake. The WWW—with its language of Universal Resource Locators, or URLs—organizes data representations of information in a highly structured way and is just one of many applications that are supported by the Internet.5

The distinction between data and information—which is directly analogous, though not identical, to the distinctions between syntax and semantics that are made in logic and linguistics—is very important in our context. While it is data that the Internet processes, and which is used in the WWW to represent information in structured ways, it is information to which human beings relate.

Indeed, the WWW is often described as being an ‘information space’, but what is the concept of a space that is being invoked here? In fact, the concept of space is quite delicate. It is in mathematics—in particular, in geometry and topology—that it has been richly developed. The idea starts with the familiar three-dimensional environment, in which everyday objects have relative position and direction, and its more-or-less intuitive generalization to the four-dimensional environment, often called ‘space-time’ in which such objects also have relative position in time.

Although a formal mathematical definition is not needed for our purposes, the mathematical concept has quite strongly influenced the informal concept that I shall need. According to the OED1:

space |speɪs|
noun [mass noun]

... Mathematics: a mathematical concept generally regarded as a set of points having some specified structure. One of the most important examples of the mathematical concept is that of topological space.

Topological spaces are a way of describing geometrical properties and spatial relationships that are unaffected by the continuous change of shape or size of figures. An important example of a topological space is metric space in which there is assigned a distance—with, essentially, the familiar intuitive meaning of physical distance—between each pair of points.

These ideas of proximity translate not only to the network architecture of the Internet, but also to the data-representation of information in the WWW (though the situation is quite complicated and here I am simplifying matters greatly).

For our present purpose, the concept of space that is useful, and which builds on the mathematical ideas mentioned earlier, derives from a key concept in computer science—namely, distributed systems (Collinson, Monahan, and Pym 2012).

According to the OED1:

distributed system |dɪsˈtrɪbjʊəd ˈsɪstəm|
noun
a number of independent computers linked by a network.
Examples of distributed systems include the following:

- The Internet itself—a vast collection of interconnected networks of computers. Individual computers connected to the Internet interact by passing messages, which they do by employing a common means of communication (which will be described later on in this chapter).
- Intranets—localized parts of the Internet that are managed by identified organizations that, typically, enforce local management and security policies to control access and use. Intranets are connected to the wider Internet by special-purpose computers called 'routers', which also employ the common means of communication.
- Cloud computing infrastructure—vast data centres, consisting of hundreds of thousands of servers, provide storage and computer services for vast quantities of data that are fed from the intranets of large numbers of clients who are logically and physically widely distributed. A key problem here, and in distributed systems more generally, is to maintain the consistency of the different copies while simultaneously maintaining the robustness of the service: see, for example, the CAP (Consistency, Availability, Partition-tolerance) theorem in the theory of distributed systems, also known as 'Brewer's theorem' (https://en.wikipedia.org/wiki/CAP_theorem), which establishes that only two of the CAP properties can be maintained simultaneously. Handling this problem is important in maintaining the experience of cyberspace.
- Mobile and ubiquitous computing (and the Internet of Things)—laptops, phones, cameras, and wearable devices such as watches and spectacles, as well as cars, domestic appliances, smart meters, and electricity substations. Indeed, almost everything on which modern society depends is integrated into distributed systems. They reside on local networks that communicate with other local networks and devices, either directly ('peer-to-peer') or via network-based servers, all employing the common means of communication.
- The global banking system—the intranets belonging to each of the world’s banks must not only provide services locally (logically locally, even if not physically locally) to the banks’ own customers but must also communicate with one other in order to support the transactions upon which the world’s commerce depends.
- Online games with multiple players—each player’s local, or home, computer runs a client copy of the game, which communicates with the central game server, which communicates with other players’ local games, and coordinates the overall interaction between all the players.

From a mathematical perspective, the basis of computer science and to which I shall return later, the key concepts of distributed systems are the following (Anderson and Pym 2016; Barwise and Seligman 1997; Caulfield and Pym 2015; Collinson, Monahan, and Pym (2012); Coulouris et al. 2011):

- **Locations**: a collection of linked places, be they physical or virtual, that constitutes the basic architecture of a system. Individual computers, file stores, and so on exist at locations within a distributed system, but locations are also the places within computers where the CPU (Central Processing Unit) and other components reside.
- **Resources**: the entities that a system uses—consumes, creates, moves—in the course of its operations. Examples of resources include the memory locations where data is stored, the
processor cycles available to perform computations, and human operators required to manage and maintain systems.

- **Processes**: the collection of activities, which are mostly concurrent, that constitutes a system’s operations, and so delivers its services. Examples of services include a bank’s customer-facing website, streaming films, and the multitude of system-level services provided by a computer’s operating system in order to perform computations, manage the keyboard and screen, manage a computer’s memory and storage, send and receive email, etc.

Additionally, a specific system, described using these components, resides within an *environment*, and resource transfers between the system and its environment characterize the service that the system provides.

So, here the relevant ‘space’ consists in the distribution of resources around the locations of the system and the presence of processes that manipulate those resources. This definition might seem a bit restrictive but, in the distributed systems metaphor, the presence and activities of a human being (using a computer, posting to Facebook, downloading a file, and so on) simply amounts to the presence of a process.

The term ‘cyberspace’ is derived from ‘cybernetics’ and ‘space’, and the meaning of the term depends essentially, though quite implicitly, on the distributed systems metaphor. The term ‘cybernetics’ was introduced in the late 1940s by Norbert Wiener (1948, 1950).

According to the OED:

> cybernetics /saɪbəˈnɛtɪks/
> plural noun [treated as sing.]
> the science of communications and automatic control systems in both machines and living things.

ORIGIN 1940s: from Greek *kubernētēs* ‘steersman’, from *kubernan* ‘to steer’.

There are several aspects of this definition that are important for the idea of cyberspace. That it refers to ‘automatic control systems’ may perhaps seem rather restrictive, but I think that should be seen as a consequence of the perspective of the age of the definition: in the 1940s, although the idea of automated control of machines was well understood, the scope of the information technological revolution that was to come had not been anticipated. It mentions also communications. As I have described, the concept of communication, and a common means of supporting it, is a key aspect of distributed systems.

So, now I have all the components, I need to return to the OED definition of cyberspace given earlier. Let us try to understand this rather concise definition using the concepts we’ve considered so far. First, what are ‘computer networks’? The appropriate metaphor here—which I have already discussed at some length—is that of distributed systems. Computers, be they servers, workstations, laptops, phones, controllers embedded in cars, aeroplanes, or refrigerators—or even entire data centres—are resources that reside at locations.

Second, what does ‘communication over’ mean? Computers residing at locations communicate with other such devices residing at other locations using wired and wireless connections. These connections transfer data between located devices using the TCP/IP protocol.
Last, what is meant by the ‘notional environment’? It seems that this is where the presence of human interpreters becomes essential. The distributed systems metaphor completely accounts for the infrastructure and its processing of data, so the ‘notional environment’ can only be something that is experienced by the users of the infrastructure.

Users provide the interpretations of data and its movement around the infrastructure that constitute the ‘information environment’. Now, in principle, every data item is a discrete entity and the collection of all such items in the (albeit vast, massively interconnected) infrastructure of the Internet is finite and so can be counted.

From the perspective of the users, however, things look very different. This is for two reasons. First, end-users (as opposed to users who are systems professionals) primarily perceive information, not data. A picture received on a phone may be a finite collection of pixels, but it represents an image of the physical world of substances and qualities. Second, the exchange of information mediates communication between humans, in a shared social space that is created by the technology and its users, and that communication is almost never wholly captured by the data that is exchanged.

Only with this last component is the definition of cyberspace—and its characterization as a mass noun—really meaningful.

The origin of the word ‘cyberspace’ does not lie in hard science. Rather, it was coined in science fiction, by William Gibson, first in a short story, ‘Burning Chrome’, in 1982 and reused a little later in his celebrated novel, *Neuromancer*, in 1984. Not only does Gibson introduce the term but he also offers a definition:

> A consensual hallucination experienced daily by billions of legitimate operators, in every nation, by children being taught mathematical concepts … A graphic representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data.

It can be seen that Gibson’s definition, albeit expressed in a novelist’s style, anticipates more-or-less all our analysis: a highly complex distributed system, the representation of data, and the presence of human minds. Indeed, it seems to capture very directly the experiences of humans who engage in ‘immersive’ or ‘virtual reality’ games with other players who may be physically located in many distributed locations, but who together inhabit a shared environment of data that they collectively, and consistently, interpret as the ‘world’ of their game.

Cyberspace is also an important component of conflict in the modern world (Rid 2013; Singer and Friedman 2014) and, consequently, the world’s military and defence agencies have considered the significance of what is increasingly known as the ‘cyber battlespace’ for their strategies and operations. Indeed, some of them have even attempted to formulate their own definitions—for example, the US Department of Defense in 2008 (Schachtman 2008):

> a global domain within the information environment consisting of the interdependent network of information technology infrastructures, including the Internet, telecommunications networks, computer systems, and embedded processors and controllers.

Here, it is important to understand that the use the term ‘domain’ refers not to what is usual in computer science, where it describes a collection of addresses within the Internet, such as everything with a ‘.com’ or ‘.uk’ suffix, but rather it refers to a domain of warfare, the other four being land, sea, air, and space. Again, it is clear that a distinguishing feature of the cyber domain is its combination of the virtual/digital and the physical.
In summary, then, what have I described about cyberspace so far?

- First, that it is a concept that builds on the physical and logical infrastructure provided by the Internet.
- Second, that, while the Internet processes data, it is information—that is, interpreted data—that is the medium of cyberspace.
- Third, that the interpretation of data, and the processing of information, are performed by humans, who are themselves essential components of cyberspace. Together, the human participants inhabit the shared social space that is an essential component of cyberspace.
- Fourth, that there is an essential interplay between—indeed, a merging of—the physical and the virtual.

So far, I have concentrated on unpacking the concepts that constitute cyberspace and their associated language. These concepts have, however, a substantial backstory through human history, and it is long and rich.

Before embarking on the story, I should note that it is not possible in a short chapter such as this one to represent fully and acknowledge all of what is a vast literature. Accordingly, the sources I reference are intended only to be suggestive of the literature, and I apologize unre- servedly to anyone who feels unjustly treated. I note also that I am not a professional historian and I make no claim to historical completeness in this article.

**Cyberspace in the Ancient and Early Modern World: Beacons and Semaphores**

Travelling between widely separated cities, by walking, riding horses, and sailing in ships takes a long time. Messages sent by these means are therefore slow to arrive. For example, during the negotiation of the Treaty of Westphalia in Münster and Osnabrück in 1648, it took two weeks for a letter to reach Stockholm. Consequently, governments and others throughout history have sought ways of communicating more rapidly.

Perhaps the simplest form of rapid long-distance communication is the beacon, a fire lighted on a hill to give warning of, say, an approaching enemy. A sequence of beacons on a chain of hills can give rapid warning over long distances—it takes just a few minutes to light a fire, and the signal then travels at the speed of light—but, of course, the language of communication is rather restricted.

In the fourth century BCE, the Greek military strategist Aineias Taktikos described a partial solution, the ‘Greek hydraulic telegraph’—as explained at, for example, https://en.wikipedia.org/wiki/Hydraulic_telegraph, where more detail and further references can be found. Fire torches were used, by the sender, to initiate and synchronize, with the receiver, a connection between operators at observation points on hills with clear lines of sight between them. Each hill had an identical container, with a valve or spigot at the bottom, filled with water and with a vertical rod floating in the water. The rods were marked with codes at points along their length. The set-up is depicted in Figure 1.1.
Once the connection was synchronized, each operator would open the valve until the water had emptied to the point marking the required code, at which point the operators would close their valves and simultaneously lower their torches. The length of time the sender’s torch remained raised determined a specific, predetermined message. In principle, such a system could be used to send messages in full written language, but, in practice, it would seem likely that the need for efficiency would dictate a small, fixed set of possible messages.

Does the technology of beacons and semaphores, as developed from the Greek world, support something that corresponds to the concept of cyberspace? First, there is an underlying physical infrastructure, chains of torch beacons and data-processing water containers, which form a network of communication routes. This network supports logical connections between the individuals wishing to communicate with other individuals at other locations. Second, while the system of beacons and containers transmit data from location to location, it is the humans who send the messages (not the operators of the
infrastructure) and who interpret the data, giving it meaning as information. Last, there is indeed an essential interplay between the physical and virtual. I think I must conclude that the fourth century BCE had a form of cyberspace.

A major advance in the development of cyberspace occurred in late seventeenth-century France. 'Le système Chappe', developed by Claude Chappe (see, for example, https://en.wikipedia.org/wiki/Claude_Chappe and also Standage’s delightful book, The Victorian Internet (1998), was a nationwide semaphore network used for (relatively complex) government and military communications. At its greatest extent, it connected Paris to Amsterdam and Calais to the north, to Mainz, Strasbourg, and Venice to the east, to Marseille, Perpignan, and Bayonne to the south, and to Nantes, Brest, and Cherbourg to the west.

Chappe’s system was a network of towers (see Figures 1.2, 1.3), each of which supported two arms that rotated into different positions. The positions included codes for letters and numbers as well as for control signals used to verify the correctness of the reproduction of the messages as they passed from one tower to the next.

Somewhat later, in 1838, an English civil engineer called Francis Whishaw also proposed a hydraulic telegraph. Whishaw’s hydraulic telegraph was based on the levels of water observed in vessels connected by water-filled pipes: a change in level at one end, representing

**Figure 1.2.** A replica of one of Chappe’s semaphore towers in Nalbach, Germany
the sent message, is reflected at the receiving end of the pipe with no perceptible time delay. This proposal represents a possible improvement in both speed and reliability, but essentially the same sense of a cyber space as the Greek version.

Amusingly, the term ‘semaphore’ persists in the modern world of information technology in the theory and practice of concurrency, in which two computer programs execute at the same time while attempting to use shared resources. In this context, a semaphore refers to a variable, or other abstract data type, that is used to control access to a resource between concurrently executing processes.

**Some Key Concepts**

Before considering cyberspace in the modern world, in the sections that follow, it is important to mention three of its other key precursors, all of which have contributed to its technological and social infrastructures:
The commercially available printing press: Printing presses have existed in China and Korea for around 1,800 years, but it is perhaps Gutenberg’s introduction of a commercially available, and well-promoted, service that marks the entry of printing into a ‘space’ of communication. Copies of single handwritten manuscripts, representing resources of knowledge and typeset using ‘standard’ characters, could be mass produced and circulated widely. Thus, knowledge could be shared around many distinct locations, commented on, modified, and further shared.

Postal services: Although commercial printing presses provided a means of mass-producing information resources, so that they might be consumed by many different individuals residing at many different locations, the realization of this sharing requires a process for circulating copies of manuscripts. Postal services provided the first reliable such processes and, in so doing, adumbrate some concepts that are important in modern cyber space. These include the following (with no particular historical period or timeline implied):

- **Addresses**: the sender of a package writes a code on the package that specifies the destination and recipient of the package; the service provider interprets the code in order to execute the process of delivering the package: for this to work, addresses must be written in an agreed, or at least recognizable, format;
- **Routing protocols**: packages might be collected from widely distributed starting points, such as letter boxes, then taken to a local collecting point and combined into large groups of packages that are moved to a distant collecting point (possibly involving many such steps), from where individual packages are delivered to their final destinations; to make this work, the provider of the postal services must implement processes that collect, sort, and distribute packages; and
- **A supporting infrastructure**: the service provider must provide the equipment (postage stamps, letter boxes, bags, vehicles, buildings, etc.) and personnel to collect, sort, and distribute the packages; the provider may also make use of other services, such as stage coaches, trains, and aeroplanes, and must agree terms of service with them.

A postal service, viewed as a service to its users, also has two key features:

- **Mass availability**: the service is available to all who are able to purchase the tokens, such as stamps, required to access the service;
- **Service guarantees**: state actors, such as monarchs or governments, might provide guarantees, with supporting policies, that packages will be delivered to their intended destinations and recipients, and that they will be undamaged in transit.

Telephone systems: Such systems also provide examples of the importance of the concepts of addresses, routing protocols, and supporting infrastructure. It also demonstrates mass availability and, at least implicitly, service guarantees.

Agar’s *The Government Machine* (2016) provides an excellent general contextual discussion for this perspective, exploring the mechanization of government work in the United Kingdom from the nineteenth to the early twenty-first century.
THE BEGINNINGS OF CYBERSPACE IN THE MODERN WORLD: SEMAPHORES AND TELEGRAPHS

A key technology of the Victorian period, the telegraph system developed in the 1830s and 1840s by Samuel Morse, provides another example of the importance of the concepts of addresses, routing protocols, and supporting infrastructure. It also demonstrates mass availability and, at least implicitly, service guarantees.

According to the OED¹:

telegraph, n. — a system of or instrument for sending messages or information to a distant place; v. — to signal (from French télégraph)

I have explained that the idea of an optical telegraph dates from the ancient world. Perhaps the most well-known early precursor to the Internet, however, and certainly the one with the strongest resemblance, is the telegraph system developed in the nineteenth century. The history of this ‘Victorian Internet’ has been elegantly and captivatingly described in Standage’s (1998) book, which, as we shall see, helps us to understand its significance for the origins of the concept of cyberspace.

Figure 1.4 illustrates the major global telegraph connections around the world in 1891. Compare with the modern map of telecommunications cables given in Figure 1.5:

![Figure 1.4. Telegraph Connections (Telegraphen Verbindungen), 1891, Stieler’s Handatlas, Plate No. 5, ‘Weltkarte in Mercator’s projection’](image-url)
First, there is an underlying network infrastructure;

Second, there are several key resources, placed around the locations of the network infrastructure, upon which the operation of the telegraph depends—namely:

- the network cables used to connect different points around the world;
- the electrical devices that generate and receive the electrical signals that are used to encode messages for transmission across the network;
- the human operators of the devices who translate between natural language and the encoded messages;
- the paper used to write down messages to be encoded and messages that have been decoded.

Third, the communication of messages between points on the telegraph network occurs as a collection of concurrent processes that utilize the resources present at locations around the network.

The telegraph system of the nineteenth century thus came very close to delivering a cyberspace. Many of the features of cyberspace that I have identified were present in the telegraph system—network infrastructure, key resources, and concurrent processes—but one aspect that was missing, at least in a sufficiently explicit form, was that of the shared social space created by the technology and its users. Although messages could be sent and received very efficiently, there was no way to post information that could be read and contributed to by other participants in the space; and there was no way to implement something like Facebook using the telegraph system.
The infrastructure of modern cyberspace

The technology supporting worldwide data communications did not significantly advance from the telegraph (wired and wireless) until the early stages of the development of what would become the Internet. Indeed, as can be seen in Figure 1.5, the pattern of connectivity even now reflects that of the wired telegraph network (Figure 1.4).

The ARPANET—which stands for Advanced Research Projects Agency Network (https://en.wikipedia.org/wiki/ARPANET), after the United States’ Advanced Research Projects Agency (https://en.wikipedia.org/wiki/DARPA) that funded its development—was the seed that would eventually grow into the Internet. It was proposed in 1968 and established in 1969, with the first link being between UCLA and Stanford University. Famously, the first message sent between the two sites was ‘LO’—the first two characters of ‘LOGIN’; the connection failed before the command could be completed.9

The ARPANET was an early ‘packet-switching’ network—in which transmitted data is grouped into blocks, called ‘packets’, that are of a suitable size (depending on things like the network’s ‘bandwidth’) for transmission across a network—that implemented the TCP/IP protocol, upon which the modern Internet depends. Packet switching (https://en.wikipedia.org/wiki/Packet_switching) stands in contrast to ‘circuit switching’ (https://en.wikipedia.org/wiki/Circuit_switching), as used in early telephone networks, in which dedicated circuits are established between two points (e.g. two telephone service subscribers) that wish to communicate. Circuit switching, which does not require the overhead of decomposing messages into packets and recomposing after transmission, could be used in the Internet. However, it makes much less efficient use of the available network capacity (or ‘bandwidth’).

Recalling our discussion of semaphore and telegraph systems, it can be seen that although they also required a notion of packet in order to send and receive messages—words are coded as delineated sequences of coded letters, Morse code (https://en.wikipedia.org/wiki/Morse_code)—they all really worked by establishing circuits between the communicating locations, as with early telephone networks. Packet switching is perhaps the key conceptual advance of the Internet over the telegraph networks.

The TCP/IP protocol is one example, a very important example, of a specification of a network communications protocol—recall from our discussion of distributed systems the essential need for ‘common means of communication’—that is tailored to the underlying physical technology that supports its operation. Such technology is not unique, however, and the Open Systems Interconnection (OSI) model provides a standardized reference—it describes the essential features of the infrastructure of the Internet. Figure 1.6 illustrates how the components of these protocols are built up, from the physical layer, providing underpinning infrastructure, through logical organizational layers, to the application layer, providing services to users. The TCP/IP model can be seen as an implementation of the OSI model—for example, the general ‘Network’ layer described in the OSI model is implemented by the Internet in the TCP/IP model.

Figure 1.6 depicts how, in a somewhat simplified and quite widely described form, the Internet can be seen as implementing cyberspace. To see this, first recall our summary of
cyberspace at the end of the first section. Note that, for the purposes of this discussion, I am taking cyberspace to be represented by the medium of the WWW and its use by humans. Does this implementation of cyberspace deliver what we expect? Recall what we learned about cyberspace in the first section:

- First, that it is a concept that builds on the physical and logical infrastructure provided by the Internet;
- Second, that, while the Internet processes data, it is information—that is, interpreted data—that is the medium of cyberspace;
- Third, that the interpretation of data, and the processing of information, are performed by humans, who are themselves essential components of cyberspace;
- Fourth, that there is an essential interplay between—indeed, a merging of—the physical and the virtual.

The first point is clearly supported by this picture: the communication between a web browser and web server, both of which have physical location as well as logical location, is implemented by a sequence of flows of data over a physical network, but that physical network supports interpretations of that data relative to a logical architecture, which organizes the information into useful systems (of knowledge, understanding, and so on). The second and third points reside in the users’ interpretations as information of the data that flows between the web browser and the web server. Finally, the fourth point summarizes the overall relationships between the components of the diagram in Figure 1.7: data and information have both physical and logical locations; data is processed at physical locations, around the loop between browser and server, but interpreted at logical locations by the human users—in, we might say, Bewusstseinslagen. Within cyberspace itself, as implemented by the Internet, the structural organization provided by the distributed systems model is not always the most helpful. Rather, it is sometimes more useful to infer information about cyberspace in terms of what the statistical
structure of the data and its flows tell us about patterns of use. In this context, topological modelling approaches such as that suggested in Ohmori and Kunii (2007) may also be helpful.

Examples of this kind of analysis include answering questions about the density of Internet use in different countries around the world (see, for example, Figure 1.8), which social networking sites are more popular in which countries, from where most phishing attacks originate, and with what levels of intensity, and so on. Figure 1.8 illustrates Internet users in 2015 as a percentage of a country’s population.
How did a world of communication based on semaphores and Morse code sent over telegraph cables become a world dependent on Internet-supported cyberspace? The answer really is the story of the development of modern computer science (though see Isaacson 2014 for a useful perspective): I cannot hope to do justice to that in this chapter. Rather, I hope to provide a conceptual framework for understanding and reasoning about cyberspace that is applicable to all these stages in the history of cyberspace.

## Modelling and Reasoning about Cyberspace

This section is primarily intended for those readers with a more mathematical, or at least philosophical, background and, in particular, for those with an interest in logic. Nevertheless, I hope that all readers who are willing to encounter a little formalism will be able to appreciate the value of the perspective I describe.

As I have described, distributed systems provide a model of computation in which information-processing devices are located on networks and communicate with one another, and with their environments, and coordinate their actions by passing messages between one another and between themselves and their environments. The resulting interaction of these components of systems and their environments delivers the systems’ services to their clients.

Mathematically, distributed systems can be described using the following concepts, as described in Section 1:

- **Locations**: Mathematically, locations are described using topological structures that give a useful account of the (physical or virtual/logical) notion of ‘place’ and ‘connections between places’. The leading example is perhaps directed graphs, but mathematically other structures can also be used (Collinson, Monahan, and Pym 2012). The concept of location, and its intended mathematical characterization in this context, provides the topological component that is a core part of the concept of space discussed earlier.

- **Resources**: Mathematically, resources are modelled by abstract algebraic structures called ‘partial monoids’. These gadgets are sets that come with an operation, which has a unit or neutral element, for combining some, but not all, of their elements (used in Collinson, Monahan, and Pym (2012) and Ishtiaq and O’Hearn [2001]). Perhaps the most important example of such a monoid is given by the set of natural numbers (with 0) less than or equal to a specified maximum, \( \max \). Combination is addition, with unit 0. The combination of two numbers \( m \) and \( n \) is defined just in the case that \( m + n \) is less than \( \max \). Another important example is given by the ‘stack’ and ‘heap’ in computer memory (RAM) (Ishtiaq and O’Hearn 2001).

- **Processes**: Mathematically, processes are described using structures called ‘transition systems’ and an important class of examples of transition systems are described by ‘process algebras’ (Anderson and Pym 2016; Collinson, Monahan, and Pym (2012); McKinsey and Tarski 1944).
The key idea is that the state of a system is described by a triple $L, R, E$ consisting in the configuration of the system's locations, $L$, the distribution of its resources, $R$, around its locations, and a description, $E$, of the processes that are currently executing. When an action occurs during that execution, the resources are manipulated, perhaps being consumed, created, or moved to new locations.

Again, as described in Section 1, systems exist within environments with which they interact (i.e. they are part of an ecosystem). This interaction is typically described in terms of the incidence of events in and out of the model.

- **Environment**: Mathematically, the incidence of actions from the environment upon a model and, conversely, the incidence of actions from a model upon the environment can be represented simply using probability distributions. Perhaps the paradigmatic example of this is the arrival of entities, be they people or packets of data, in a queue, where the arrivals at the queue are described using the negative exponential distribution (Ross 2014), which has just one parameter, the 'arrival rate'.

With this machinery in place, I have, mathematically speaking, all I need to describe the logical and physical architecture of the Internet—that is, the infrastructure of cyberspace:

- **Locations** in the Internet are given by a range of examples that are relevant for my discussion:
  - The physical network graph: Figures 1.3, 1.4, and 1.5 give examples of network graphs; the global scale network connects regional and national networks, which, in turn, connect organizational and domestic networks.
  - The virtual network graph: organizations may be distributed across the world and yet appear to be a single network location—for example, a multinational corporation may have physical presence in many countries, but its networks may all be part of the same family of IP addresses, so that they appear as part of the same network location even though they are in many different physical locations.\(^{12}\)
  - The locations of the human users of the Internet, participants in cyberspace, and the devices with which they interact and upon which the services they use depend.

- **Resources** in the Internet are things like computers, some providing computation, some providing network management, and some providing storage; peripheral devices, such as printers and scanners; security devices such as IDSs and IPSs;\(^{13}\) and people, such as programmers, system administrators, and end users (the participants in cyberspace).

- **Processes** in the Internet are the things that happen. For one example, an individual computer's operating system is a program that executes continuously in order to provide all the computer's services to its users: screen, keyboard, network connection, application execution, and so on; for another, the services provided by the network of servers and routers that support the operations of an Internet service provider; and, for another, the human resources, financial, and other business processes followed by the users of information and management systems.
And, finally:

*Environment* in this context provides a way for the modeller to focus on a particular part of the Internet, or indeed of cyberspace, while retaining an appropriate representation of the rest of the network on that specific part. While the part of interest is modelled in detail, using the concepts of location, resource, and process as described earlier, the interaction of that part with the rest of the network is modelled simply in terms of the incidence of events across the boundary of the part modelled in detail.

Although I have described a framework for describing the underlying conceptual and technical infrastructure of cyberspace, I have not yet provided a way to describe cyberspace itself: even if we consider that humans and their interaction with the architecture, and, indeed, other humans, can be described using located resources and processes, we still lack a natural way to talk about the interpretation of data and how humans reason about it. For that, we are going to take our final step: to logic.

Logic is the science of reasoning. It is studied within computer science, mathematics, and philosophy. It means the same thing in all these areas, although they each tend to emphasize different aspects of its study. They interact with one another very fruitfully. In computer science, in particular, it is important to become accustomed to the idea that there is no single, all-encompassing system of logic that is well adapted to all the different kinds of reasoning that are needed.14,15 The discussion in Pym (2019) of the use of logic as a modelling technology may be useful for some readers.

Mathematical models of distributed systems of this kind are closely associated with ideas from logic. Logic is the science of reasoning and one of its key ideas is that of truth. In logic, truth is a very precisely defined concept. It relies on a few key ideas: syntax, semantics, and interpretation. Syntactic entities are interpreted as semantic entities, just as data is interpreted as information. Truth is a property of a logical formula (which is a syntactic entity) relative to a *model*. A model is a mathematical structure that describes relationships between semantic entities.

In modal logic (Blackburn, de Rijke, and Venema (2001), which is perhaps the key tool in the logician's kit for reasoning about action, the ideas of necessity and possibility, so-called 'modalities', can be expressed. The key to understanding these ideas comes from some beautiful work initiated by Saul Kripke (1963), the application of which in systems modelling is discussed in, for example, Anderson and Pym (2016); Caulfield and Pym (2015); Collinson, Monahan, and Pym (2012); Simon (1996).

The main idea is that truth is defined relative to a *world*. The concept of a world is philosophically quite delicate, but for our present purpose we can think of it as a state of knowledge or the state of a system. A collection of such 'possible worlds' that might be taken as a place to give meaning to formal logical expressions can be seen as a space in the sense that we have already discussed. In fact, this kind of semantics can be formulated explicitly in terms of topological spaces, which are perhaps the prototypical mathematical example of the concept of a space (Tarski 1969; McKinsey and Tarski 1944).

What is most important for my story here is that the set of all worlds, $W$, must come with a partial ordering16 on its set of elements, so that we define truth relative to models $M$ of the form $(W, \leq)$. 
Given a world $w$ in the set $W$,

$$w \models_M \phi$$

denotes that the logical formula $\phi$ is ‘true in the state $w$. For example, if $w$ is a state in which there are precisely three apples and two oranges, then the formula $\text{More (Apples, Oranges)}$, which is intended to mean that there are more apples than there are oranges, is true at $w$. But, if $w$ is any state in which there are at least as many oranges as apples, then the formula is not true there.

Suppose now that our states are ordered as follows: $w \leq v$ just in the case that $v$ has more apples than $w$.

In general, $w = M \square \phi$ denotes that the formula $\phi$ is necessarily true at the state $w$ in the model $M$. This is defined as follows: $w \models_M \square \phi$ just in the case that, for every state $v$ such that $w \leq v$, it is the case that $v \models_M \phi$. So, in our little example of apples and oranges, the formula $\square \text{More (Apples, Oranges)}$ is true if $w$ is a state in which there are precisely three apples and two oranges because any state that is beyond $w$ must have more apples than oranges. Note that we may choose to consider many possible models. Different models will, in general, have different sets of possible worlds.

Similarly,

$$v \models_M \Diamond \phi$$

denotes that the formula $\phi$ is possibly true at the state $w$ in the model $M$. This is defined as follows:

$$w \models_M \Diamond \phi$$ just in the case that, for some state $v$ such that $w \leq v$, it is the case that

$$v \models_M \phi$$

So, now supposing that states are ordered so that $w \leq v$ just in the case that either $v$ has more apples than $w$ or $v$ has more oranges than $w$, then there is state beyond $w$ at which

$\Diamond \text{More (Oranges, Apples)}$

is true, so that

$\Diamond \Diamond \text{More (Oranges, Apples)}$

is true at $w$.

What has logic got to do with cyberspace? Just about everything, actually, if one believes that logic provides a good, or at least useful, account of human reasoning. Moreover, computers are inherently machines that implement logic. Assuming that at least, then logic provides the essential link between people and the systems that support cyberspace, as described in Figure 1.9.

The diagram indicates the relationship between reasoning about systems (including about other humans within the system) and logical reasoning about mathematical models of the system. The ideal situation is when this diagram ‘commutes’—that is, for a given system, the logical formalization of human reasoning about its properties corresponds exactly to logical reasoning about a formal model of the system. Such a situation is very rare indeed, and really
only works out in the context of very specific reasoning tools (see, for example, discussions in Apt, de Boer, and Olderog 2010; Ishtiaq and O’Hearn 2001; Caulfield and Pym 2015; Pym, Spring, and O’Hearn 2018; and Simon 1996).

I have already given an example of a model in the context of distributed systems. It is built out of descriptions of the system’s locations, resources, and processes. Triples of locations, resources, and processes are the states of the model of a system. The ordering of the states is then given by the evolution of the model as actions occur (Anderson and Pym 2016; Collinson, Monahan, and Pym 2012; and Milner 1989).

Actions are the basic building blocks of processes, one of the core components in our model of the infrastructure of cyberspace. When actions occur, the state of a system changes. For one example, a computer program may perform a ‘write’ action to put value, a number representing a resource, to a memory, location. For another example, a human user of a computer may give the ‘print’ command, so causing data to be copied from the computer’s memory to the printer, followed by the consumption on ink and paper and resources, and the creation of a document.

Thus we have $L, R, E \leq M, S, F$ just in the case that $L, R, E$ can evolve to become $M, S, F$ by some action. The notion of logical truth supported by such a mode, written as $L, R, E \models_{M} \phi$, is read as ‘the process $E$, executing with respect to resources $R$, at location $L$, has property $\phi$’.

Then we can define versions of the necessity and possibility modalities that are parametrized by actions. The counterpart to the necessity modality is $[a]\phi$, which is read as ‘the formula $\phi$ is necessarily true after the action $a$’. Similarly, the counterpart to the possibility modality is $\langle a \rangle \phi$, which is read as ‘the formula $\phi$ is possibly true after the action $a$’. More formally, we define them as follows:

- $L, R, E \models_{M} [a] \phi$ holds just in the case that, for every evolution of $L, R, E$ to $M, S, F$ by the action $a$, we have that $M, S, F \models_{M} \phi$ holds;
- $L, R, E \models_{M} \langle a \rangle \phi$ holds just in the case that, for some evolution of $L, R, E$ to $M, S, F$ by the action $a$, we have that $M, S, F \models_{M} \phi$ holds.

**Figure 1.9.** Reasoning about systems and reasoning about system models
These definitions explain how logical reasoning about (the data held by) systems interacts with the actions performed by the system. In particular, they begin to explain, in terms of information processing, how humans interact with the system and other humans.

In order to see an example of all this, think about the picture of the implementation of cyberspace given in Figure 1.6. Suppose a user, sitting at a computer in their home, is using a web browser to access an online store (let us call it ‘BigRiver’). The user is looking at the webpage for Oxford University Press’s Handbook of Cybersecurity and clicks on the ‘Buy now’ button. If the user’s bank account has sufficient funds and if the BigRiver website has given the correct information about the availability of the book, then, provided everything works as it should, the book will be sent to the user. We can describe this situation logically as follows:

\[ \text{home\_computer, bank\_account, BigRiver\_website} \models \text{Cyberspace (buy\_now) book\_sent} \]

That is, located at their home computer, with the resources available in their bank account, while running the BigRiver website process, the user may click on the ‘Buy now’ link and it is possible, if all goes well, that the book will be sent to their specified address.

Note that this logical assertion describes a state of affairs in the part of the diagram described as ‘Cyberspace’. We could use similar logical assertions to describe the (many) states of affairs that must obtain in other parts of the diagram in order for the assertion about Cyberspace to be realized.

At this point, the reader might be forgiven for thinking that the logical language that explains all this is rather impoverished. I would agree, although I would note that, in fact, the framework I have suggested, as developed, for example, in Anderson and Pym 2016; Apt, de Boer, and Olderog 2010; Caulfield and Pym 2015; Collinson, Monahan, and Pym 2012; and Milner 1989, can express a great deal, albeit somewhat tediously. A challenge for the community of logicians, if logic is to demonstrate what I believe is its full potential as a modelling technology, is to develop concise and powerful representations of logical reasoning.

In fact, the world of (modal) logics for reasoning about actions (performed by agents) is much richer than I have so far suggested and includes epistemic logics, for reasoning about agents’ knowledge of systems, doxastic logics, for reasoning about agents’ beliefs, and temporal logics, which incorporate a representation of time (system events, such as the sending and receiving of messages, which occur at relative points in time). All these systems of logic build on the basic ideas sketched earlier. The Stanford Encyclopaedia of Philosophy (https://plato.stanford.edu) provides a great deal of information about these systems of logic, and there is a vast literature in computer science that is concerned with their use in reasoning about systems.

Perhaps the most important of these many possibilities for our story are the epistemic and doxastic logics. Roughly speaking, in these logics, modalities are parametrized not by actions but rather by agents, who of course can perform actions. Agents may be humans or system processes, and epistemic and doxastic logics (again, see the Stanford Encylopedia) allow us to reason about their knowledge and beliefs. Exploring the use of these logics, and their relationships with tools from behavioural economics (e.g. Baltag and Renne 2016), game theory (e.g. Binmore 2007), and psychology (e.g. Kirchler and Hoelzel 2017), to reason about the behaviour of agents in cyberspace would be another chapter in exploring the origins of cyberspace.
Of course, individuals and organizations do not always behave ‘logically’ when they interact with one another and with the systems that support cyberspace. They behave in ways that others may consider to be irrational. This observation does not really undermine the perspective presented here. It has very little to do with logic in the sense that I have described, which is about the mechanism by which conclusions are drawn from chosen assumptions. Some assumptions may lead to what may be described as ‘irrational’ behaviour, even though the logical mechanisms may be perfectly sound.

**Summary**

I have sought to explain the origins of ‘cyberspace’ historically, linguistically, and conceptually. I have explained how the idea of cyberspace derives from a complex combination of physical and logical structure, which supports complex interactions between and among humans and information-processing machines, and I have given a conceptual and mathematical framework for modelling the conceptual and technical infrastructure of cyberspace.

I have also explained how logic can provide tools, based on our approach to modelling the infrastructure of cyberspace, for capturing how humans and other agents reason about cyberspace, and so, to some extent at least, how they experience cyberspace. Experience, of course, involves more than logical reasoning alone. Exploring that dimension would be yet another chapter in exploring the origins of cyberspace.

I have explained how the essential features of cyberspace have been part of the human experience, ‘a consensual hallucination experienced daily by billions of legitimate operators’ as Gibson (1982) put it, for a very long time and, as the science fiction writers continue to predict, we can expect that there is much more to come. I have not discussed questions of security in cyberspace—that is the topic of the rest of this handbook. Again, Standage’s 1998 book (Chapter 7, ‘Codes, Hackers and Cheats’) provides a delightful starting point.

**Acknowledgements**

I am warmly grateful to Tristan Caulfield and Jonathan Spring for their thorough and thoughtful advice on drafts of this article.

**Notes**

3. It is not my purpose here to explore the definition of ‘information.’ The literature on the subject is substantial. For our purposes, the usual understanding of a generally well-educated reader—see, for example, the definition provided in the OED—will suffice. For a philosophically sophisticated discussion that is beyond the scope of this article, see, for

   OSPF: https://en.wikipedia.org/wiki/Open_Shortest_Path_First

5. The distinction between the WWW and the underlying Internet is an essential one, but there would be no WWW in its current form without the underlying communications architecture. Some of the key concepts of the WWW—for example, the concept of hypertext and the underlying ideas of distributed systems (Coulouris et al. 2011; Simon 1996)—have, as described in Berners-Lee's research proposal (1989, 1900) and book (2000), prior histories in computer science. The linking of hypertext to the Internet through 'http' (HyperText Transfer Protocol) and URLs and HTML (HyperText Markup Language) are Berners-Lee's definitive contributions.

6. The Treaty was a huge diplomatic effort: 176 negotiating teams representing 194 agents; the French brought a 150-strong entourage, including pastry chefs, priest confessors and dancing instructors. The negotiators had constantly to check with their capital: a letter took 2 weeks to reach Stockholm, 4 weeks to reach Madrid: R. Boyes, 'Treaty that created 'the soil of despair', The Times, 24 October 1998. I am grateful to Paul Cornish for this reference.

7. 'Telegraph' means 'distance writing' in Greek.


9. The SAGE (Semi-Automatic Ground Environment) missile defence system, which was developed by MIT’s Lincoln Laboratory and which operated in the United States from the late 1950s to the 1980s, is also a seed (https://en.wikipedia.org/wiki/Semi-Automatic_Ground_Environment).

10. For our present purposes, I refrain from considering AI alternatives to humans as interpreters of data to yield information.

11. German: a state of consciousness or a feeling devoid of sensory components (Merriam Webster).

12. An IP address is a numerical label (a sequence of numbers representing a 32-bit or 128-bit number) assigned to each device connected to a network that uses the Internet protocol (IP) for its communications. An organization may, for example, own all of the IP addresses that begin with a given sequence of numbers.


14. That is not to say that many things expressible in one kind of logic cannot be expressed in another; rather, that it may not be convenient to do so.

15. In addition to the classical propositional and predicate logic that is routinely taught to undergraduates in computer science, mathematics, and philosophy, we can add, among other things, modal, temporal, and epistemic systems, and their higher-order, intuitionistic, and substructural variants. The many volumes of the *Handbook of Philosophical Logic* and the *Stanford Encyclopaedia of Philosophy* (https://plato.stanford.edu) provide starting points for exploring these topics.

16. A partially ordered set formalizes the intuitive concept of an ordering of the elements of a set. A ‘partial order’ on the set relates pairs of elements of the set in such a way that the
relationship been the elements of the pair is reflexive (every element is related to itself),
anti-symmetric (no two elements precede each other in the ordering), and transitive (if
the higher of one pair is below the lower of another pair, then the lower of the former
pair is below than the higher of the latter pair). Not all pairs of elements of the set need be
related by the order; such pairs are ‘incomparable’ in such an order.

17. Technical note: for an elegant explanation of the structure of processes, including
concurrency, non-determinism, and recursion, see Robin Milner’s Communication and
Concurrency.

Bibliography

Dordrecht: Springer.
dynamic-epistemic/
Cambridge: Cambridge University Press.
Archive, https://www.w3.org/History/1989/proposal.html
University Press.
London: College Publications.
Harlow: Pearson.
https://plato.stanford.edu/entries/logic-modal/
Isaacson, W. 2014. The Innovators: How a Group of Inventors, Geniuses, and Geeks Created the
Ishiaq S., and P. O’Hearn. 2001. ‘BI as an assertion language for mutable data structures’,
Proceedings of the 28th ACM SIGPLAN-SIGACT Symposium on Principles of Programming


CHAPTER 2

OPPORTUNITY, THREAT, AND DEPENDENCY IN THE SOCIAL INFOSPHERE

GREG AUSTIN

INTRODUCTION

The previous chapter described the revolutionary emergence of a suite of combined technologies (information and communications) as a powerful enabling force for many areas of human activity. These technologies and their exploitation have created a whole new ecosystem of human endeavour. This chapter looks at the social constructions of that information ecosystem—the infosphere. Using very broad sweeps, the chapter looks first at pre-political conceptualizations of cyber space as a moral sphere by scholars, technologists, and futurists. This overview sets the scene for a brief discussion of opportunities and threats that have shaped security in cyber space. The final section of the chapter offers an original analysis of the issue of dependency, which may be one of the more under-appreciated characteristics of the cyber age.

Terminology in this field remains problematic. To many readers, the term ‘cyber security’ implies a focus on protection of machines, systems, networks, the information they hold, and the communications systems that propagate the information. By contrast, the concept of ‘security in cyber space’ opens up a broader perspective, and implies consideration of social impacts arising outside the technological and physical perimeters of cyber space. The main social characteristic of defence of the new domain may not be the security of the zeros and ones that constitute the data, but protecting the political, economic, commercial, and personal power that results from new technological possibilities.

We can even go much further, as many do, and see ‘knowledge’ as the higher power compared with ‘information.’ Security of the knowledge ecosystem that depends on cyber space may be a far broader concept yet. These wider views lead us well beyond the idea of cyber security as a professional occupation, encompassing people with advanced technical skills, to a view of security in cyber space as the far bigger challenge—politically,
economically, and militarily—of securing the information society, the knowledge economy, and our digital daily life.

The idea of security for the emerging information ecosystem finds endorsement in annual resolutions of the United Nations (UN) General Assembly, beginning in 1998, on ‘Developments in the field of information and telecommunications in the context of international security’. It also finds strong endorsement by states beginning in 2002, when the UN convened the World Summit on the Information Society. This process saw the information society as global in scope and revolutionary in its impact on state power. These characteristics imply a new global obligation to protect a society ‘where everyone can create, access, utilize and share information and knowledge’ to ‘achieve their full potential’ in ‘improving their quality of life’ (United Nations 2003).

Is the Information Age a Revolution?

If the information ecosystem is a social order, then it follows that this order itself is valuable and demanding of protection independently of the machines, networks, and other technologies on which it depends. As the ancient philosopher Plato reminds us in The Republic, the security of exchange provides the primary reason for forming a society and constituting a state. The same concept exists in the Eastern tradition: ‘only when families are regulated are states well governed; only when states are well governed is there peace in the world’ (Confucius, The Analects). In any social order, stakeholders compete for position and for defence of norms that protect their positions. Norms, security and enforcers, heroes and villains, costs and benefits, winners and losers, are a part of any social order.

Plato and Confucius were both religious. In the modern era, we must understand that the social order of the infosphere, and its security, can still have an idealist or even religious orientation, or even an atheistic or materialist orientation.

In addition to being a type of social order that inherently creates, or at least demands, security relationships, the infosphere has another quality that is especially important when we consider what those security relationships might involve. This is the relationship between information, knowledge, and power that is captured in the twin adages: ‘information is power’ and ‘knowledge is power’. Philosophical sources that have contemplated the relationship of knowledge and power before the modern era are numerous, and include sources from Asian writers and folk wisdom before the rise of Greek and Roman philosophy in Europe. Thomas Hobbes, in his original English version of Leviathan (1656), saw the sciences as ‘small powers’ but, when the Latin version was published in 1668, it omitted the word ‘small’ to make the phrase ‘scientia, potentia est’ (subsequently taken as meaning ‘knowledge is power’) (Bell 1973). If these thinkers are right, and in my view history proves them to be correct, then security in the information age involves political dimensions of state power domestically and internationally, the underpinnings of corporate power and wealth, and the quality of citizen empowerment and independence.

It should be noted though that these earlier versions of the concept that knowledge is power saw it as only one of many forms of power. Hobbes is particularly relevant for this, because he included a list of sources of power (affability, prudence, nobility, friendship, and
even reputation for power). The question posed by the information age is whether by its character it assigns knowledge as the dominant form or source of power relative to others. This is addressed later in this section.

Because knowledge (synthesized and processed information) delivers political and social power, ethical (values-based) framings are both possible and inevitable in understanding the security of the infosphere. Any security concept for the infosphere can and must include active defence of the values and norms that any social group attaches to the infosphere, whether these relate to the common needs of humanity, the often-conflicting self-interest of states, business imperatives, or personal privacy. Moreover, any practical concept of security in cyber space will be dependent on at least seven quite distinct contexts: privacy and civil surveillance, freedom from harassment, resilience of the digital economy and individual businesses, protection against cyber crime, state-on-state espionage, political subversion (including activity related to globally distributed terrorism), and cyber-enabled warfare. These contexts, and others, are addressed in subsequent chapters.

The first challenge in approaching the security impacts of the information age is to answer the question of whether it is truly transformational of almost everything, as was the industrial revolution, or whether it is merely one additional aspect of a modern technological society in which information and communications technologies (ICTs) sit alongside other enabling technologies. Building on his 500-page book on post-industrial society written in 1973, Bell later used a number of shorter formats to predict, correctly it seems, that knowledge exchange would become the primary foundation of all economic and social exchange in a world that had become an information society (Bell 1976).

In a rather prescient insight, Bell also warned that the change in scale provided by the information age would of necessity dictate a change in institutions. He said a ‘major political problem’ in the post-industrial society will be its information policy. He mused on whether one of the new institutions would be an information utility (perhaps as Facebook, Google, and Twitter subsequently emerged). He then predicted that the ‘politics of information handling will arise from the nature of information itself’. As I see it, this is precisely the situation the world is confronting as it struggles with problems like political hacking and so-called ‘fake news’.

One of the most insightful analysts, however, was a Japanese technologist, Yoneji Masuda. He gave us the term ‘computopia’, in this way anticipating comprehensive and profound effects of ICTs and their use (Masuda 1981). He foresaw ‘the realisation of a society that brings about a general flourishing state of human intellectual creativity, instead of affluent material consumption’. He anticipated e-democracy, the globalization of a new renaissance, a shattering of previous conceptions of privacy (a ‘Copernican turn’), the emergence of a new concept of time value (meaning a paradigmatic shift in economic and social valuations of time spent on a task or activity), and a new intensity in system innovation—all premised on the complete objectification and commodification of information.

There are scholars who take a less transformationalist view. Frank Webster’s *Theories of the Information Society* challenges the idea of the information society as a dominant new paradigm supplanting other modes of social relationships. In his chapter, ‘What is an Information Society?’, he portrays the information society as a means of instrumentalizing certain novel aspects of social change without redefining human reality (Webster 2006). He leans closer to the idea of an informatization of life that stems from the continuity of established forces.
Robert Hassan’s *The Information Society* is critical of both camps: the ‘booster’s’ and the critics of the information society idea. He focuses his gaze on the conjunction between neoliberal globalization (capitalism) and the ‘revolution in the development and application of computer-based technologies’ (Hassan 2008). He suggests that the information society has created not just ‘pathways of possibility’ but also a ‘democratic vacuum’ that will need to find a new controlling impulse from within, or ‘we will continue to accelerate towards destinations unknown’. Manuel Castells (2012), in *Networks of Outrage and Hope*, argues that politics has been forever transformed in certain dimensions, for better or for worse, by the new information technologies and their mass application, especially in social media forms. He foresees ‘the uncertainty of an uncharted process of political change’. Established institutions may hold on to power, especially if they can co-opt the more popular themes of activists, but the more citizens can convey their messages, the more their consciousness is raised. It also carries the implication, he says, that ‘the more the public sphere becomes a contested terrain, ... the lesser will be the politicians’ capacity to integrate demands and claims with mere cosmetic adjustments’.

Against this background of divergent views, a framework offered by Floridi and Sanders (2002) is particularly useful in laying out four possible discrete approaches to characterizing the information society, and these can be used to understand the security dimension:

- a professional approach, seeing the field of information technology as something akin to medicine or law, and therefore demanding a set of professionally bounded information ethics;
- a radical approach, seeing the information age as transformative and novel—‘absolutely unique issues, in need of a unique approach’;
- a conservative approach—we only need a ‘particular applied ethics, discussing new species of traditional moral issues’;
- an innovative approach (one step down from the radical approach) that ‘can expand the metaethical discourse with a substantially new perspective’ (p. 2).

Floridi and Sanders opt for the ‘innovative approach’ and conclude that the ethical issues raised by the information age ‘are not uncontroversially unique’ but ‘are sufficiently novel to render inadequate the adoption of standard macroethics’. While that is a philosophical and ethical analysis, it is directly useful for policy. Floridi (2005) sees a practical application of his ethics: it ‘provides the conceptual grounds that then guide problem-solving procedures’. One implication of his approach is that, while not everything changes fundamentally, all in policy is affected sufficiently by the information society that it ought to be re-evaluated from that perspective.

In later works, Floridi appears to go further and approaches the transformationalist view. He suggests that the world is undergoing a transition of its mode of existence and survival every bit as radical as the transition from prehistory to history (Floridi 2012). He says that it is detaching us from future generations. Floridi predicts that the world will become ‘increasingly synchronised (time), delocalised (space) and correlated (interactions)’ (Floridi 2013). He warned against seeing this as the ‘friendly face of globalization’, suggesting that we should have no expectation about the degree to which this new world would be widespread or inclusive. In a later chapter, he posits a profound impact (dislocation) on our values and ethics: ‘Innovative forms of agency are becoming possible; new values are developing and
old ones are being reshaped; cultural and moral assumptions are ever more likely to come into contact if not into conflict. This process carries a risk of fragmentation of group values. As Paul Cornish observes on this point, ‘ethics is profoundly inter-personal, and might therefore become increasingly difficult as the information society atomises to the individual level’. At the most abstract (philosophical) level, this single unifying characteristic has been described as ‘distributed morality’. In social science terms, as opposed to moral philosophy, this might be described as distributed authority. The mass availability of computational systems and universal, ubiquitous, and instantaneous access to information carries with it a diffusion of power to a new multitude of agents.

Floridi (2013) offers four imperatives for responding the infosphere from a moral perspective, and I have adapted them to the policy domain as follows:

- accept its political imperatives (assign a social value to information and how it is used, as well as to the machines that produce and disseminate it)
- grasp the centrality of distributed political morality (pre-existing moral and political authority becomes disaggregated, and a reaggregated collective responsibility emerges);
- recognize that the enablers of information dissemination and reaggregation on a special political significance;
- regard the welfare of information itself as a central reference point of policy.

This leads to an argument that ICTs work as powerful political enablers and help distributed authority to emerge. Noting that moral enablers for the information age do not need to be only ICT-based, Floridi suggests that ‘Agents (including most importantly the State) are better agents insofar as they not only take advantage of, but also foster the right kind of moral facilitation properly geared to the right kind of distributed morality’. He says that this elaboration by information ethics helps us to realize that ‘in ethics [generally], moral facilitation is a much more influential, macroscopic and perhaps necessary phenomenon—not merely limited to computer ethics contexts—than we suspected in the past’.

To summarize, the philosophy of information ethics suggests the following useful reference points for analysing both the ethical and political aspects of an information society:

- The information age is ethically (and politically) transformative, for better or for worse.
- The new ethics (and new politics) flow from the character of the infosphere (informational ecosystem).
- In the infosphere, all informational objects have a potential moral (and a political) value.
- A moral (and politically desirable) action can be one that contributes to the flourishing of the infosphere without doing harm to any aspect of it.
- Making a moral (and politically desirable) difference may depend on the aggregation of individual information acts.
- Distributed morality (and distributed political authority) may be the hallmark of information ethics (and political life): we participate in an infosphere, a social milieu in which the diversity of sources of potential moral authority is unprecedented.
- ICTs work as moral enablers in the process of distributing morality (and political authority) away from the government to the citizen.
- The process of moral facilitation is central to the health of the infosphere.
One inescapable conclusion from this broad philosophical perspective is that cyber security as a professionally bounded and technologically oriented set of concepts, research, technologies, and practices cannot be separated easily from broader conceptions of the security in cyber space of individual people, their community groups, the businesses that underpin their daily life, their national governments, and a globalized infosphere.

**Transformational Security: For Better and For Worse**

We can understand the opportunity presented by the information age in many ways, most of them now being familiar elements in daily discourse, as well as continuing sources of wonder: from the Internet, the world wide web, and the role of social media in the Arab spring, to driverless cars, robotics, laser surgery, nanotechnologies, and exploration of deep space. Speed, ubiquity, and the global reach of ICTs now support and improve daily existence in ways previously imagined only by science fiction writers.

What is the common social element of all these amazing innovations that enhance and enrich human existence? It is the empowerment of billions of people to do old things better and faster, or in completely new ways, and to create new things to do on a vast scale. This is distributed (disaggregated) power, which brings with it a distribution or dilution of pre-existing social political power.

Alongside the opportunities, the threats are as daunting. As Floridi, Hassan, and Castells make plain, the information revolution does not have an inevitably positive moral character. The dangerous side of the new age has been seen in mass surveillance for suppression of political dissent, development of a cyber arms race, cyber attacks across international borders, cyber bullying of children ending in suicide, large-scale cyber crimes, and cyber espionage for political influencing in democratic elections. Some of the world’s leading scientists and technologists, such as Bill Gates and Steven Hawking, have warned of a threat to the very existence of humanity from advanced artificial intelligence (Cellan-Jones 2014). The UN is scrambling to agree controls on lethal autonomous weapons systems, even as primitive versions are already being deployed by major states in limited roles, such as perimeter defence of Russian nuclear missile sites or in internal security surveillance in China.

The common social element of all these information innovations that diminish and degrade human existence is the novel form of reaggregation of the distributed (disaggregated) power. It is the re-concentration of social and political power outside pre-existing institutions. This reaggregation creates a reordering of previously assured social realities facilitated by new ICTs. The reaggregation has created new ways of hurting people, society, and a country’s national interests as much as it has enhanced them. As an example, we might cite the emergence of the Donald Trump campaign and presidency as a new institution independent of and outside the Republican Party as a result of information age phenomena: the combined effect of Facebook, Twitter, and traditional media (such as CNN), aggravated by the manipulation of leaked information from intelligence agencies, both Russian and American.
If disaggregation and subsequent reaggregation of authority and power are in fact common elements to both the good and bad sides of the information age, what then is the practical significance of this insight for security in cyber space? It means that we must better understand not only the role of the enablers and facilitators (adapting Floridi’s terminology) but also the processes that correspond uniquely to specific types of information that the enablers and facilitators are trying to leverage to create new social outcomes that affect our security. This dictates a novel requirement to study two stages of this new moral facilitation: first, what processes work most powerfully to support political and social exploitation of novel types of information; and, second, how skilful are particular enablers or facilitators in mastering those processes.

**Dependence as an Outcome of Information Enablement**

One important real-world outcome from new forms of information enablement may be that of dependence created between the users of reaggregated information (in a physical product or a social communication) and the enablers. This means that the potential of ICTs to achieve new things can only be realized through the creation of a new state of dependence for the user upon the unrevealed processes underpinning the product or communication. In simple terms, dependence is, for social actors, a state of reliance on something or someone. In the material (non-social) world of objects, such as chemicals, solar systems, or electronic ICT signals, dependence is a relationship between an outcome and a necessary precondition. New forms of dependence affect the human operation and exploitation of ICTs at all levels.

The previous paragraph assumes a common sense meaning to the word ‘dependence’, which is often used interchangeably with ‘dependency’. There is a more complex approach, albeit slightly esoteric, that adds depth to our understanding of the character of dependency in the information age. Caporaso (1978), writing in the 1970s about this topic in international political economy, felt that the generic term ‘dependence’ served little useful purpose. He called out a difference between the use of ‘dependency’ as a term meaning absence of actor autonomy and the term ‘dependence’ used in the sense of a highly asymmetric form of interdependence. He implied mutual control, while the latter implies self-control. He arrives later at the important observation that it may be ‘impossible to reduce dependency to a single, unidimensional concept’. He suggested that it might be best viewed as a ‘synoptic term for a body of theory’ that addresses qualitatively different dependence types. This overarching approach also allows a necessary complementary lens, in addition to dependence and dependency. This lens is less about objective analysis of either of those two conditions and much more about knowledge of the dynamic system in which the dependence or dependency exists. For political economy, this was in his view, ‘integration of knowledge about the state and private sector within a single country which is then analysed with an external environment’.
Caporaso’s ontology is directly applicable to dependence and dependency in the information age, and therefore to security in cyber space. The need for a ‘whole of system approach’ to understanding dependence in political economy finds reflection in the view of Hathaway and Klimburg (2012) that a ‘whole of system approach rather than a whole of nation approach is the more mature form of cybersecurity’. Returning to this theme in another paper, Klimburg and Healey (2012) observe that the ‘complicated international dimension of cyber security has not always been fully understood’. Cornish suggests that the growing interdependencies ‘should defeat any notion that cyber-security is divisible: between foreign and domestic; between military and civil; and between governments and other intergovernmental or indeed non-governmental actors’ (Cornish 2009).

Here are four examples of dependence that redefine security in the information age at the individual, community, national, and international levels. These examples are meant to be only illustrative of the very large number of cyber space dependencies, billions upon billions ($10^n$) of which that might be catalogued. All new risk arising in cyber space at any level can be expressed as a function of new dependencies.

Example 1: A single person is newly dependent and disempowered through committing large amounts of personal information into the care and trust of new ICT machines, their operators, and their users in circumstances (often in foreign countries) where the individual has also surrendered (or never had) the power to revoke access to, or re-privatize, the personal information.

Example 2: A community (or a corporation) locks itself into complex dependency when it agrees to computerized management of essential services such as water supply and electricity where there is no mapping or public disclosure of the parameters that will lead to interruption of supply, and where resilience strategies have not been put in place to compensate for sudden and/or sustained loss of authorized computerized control.

Example 3: A country has extreme dependency on ICTs when the overwhelming share of its public services (in finance, health, transportation, food supply, policing, and military defence) are delivered or coordinated through tens of thousands of unique computerized systems.

Example 4: The security or prosperity of groups of states (the international level) have a high dependency on ICTs when their critical services, or their armed forces, are built on physical systems that depend on global supply chains involving several countries for delivery of secure and assurable operation of the systems.

These dependencies exist regardless of whether or not the computerized systems involved have high degrees of inherent security. The dependency arises from the information character of the social exchange resulting from the technologies, but not from the technologies themselves. Dependency is not alleviated by better cyber security of the systems. Risk of certain types of threats arising from the dependency may be alleviated by better cyber security but systemic dependencies cannot be attenuated in that way. As Rinaldi et al. (2001) note, higher degrees of complexity in ICTs increase the degree of dependency and, as a result, ‘more complex, and more extensive interdependencies lead to increased risks [and] greater requirements for security’.

That said, the inherent insecurity of many computerized systems and their vulnerability to attack become aggravating factors that exploit the dependency but, in social and political terms, the dependency is a different and pre-existing phenomenon from the phenomenon of vulnerability even if, as Cornish notes: ‘where there is dependence there is also vulnerability’
(Cornish 2011). For security in cyber space, dependency must be mapped independently of assessments of vulnerability of threat. This is essential to develop in advance some sort of mitigation: ‘where interconnectedness and dependency are not managed and mitigated by some form of security procedure, reversionary mode or redundancy system, then the result can only be a complex and vitally important communications system which is nevertheless vulnerable to information theft, financial electronic crime, malicious attack or infrastructure breakdown’.

The list of dependencies studied by moral philosophy and science, including social science, is long. It includes topics like religion dependence, drug dependence, gender dependence, oil dependence, or even more complex structural approaches like co-dependence. In respect of the information age, the term ‘dependency’ is used in several different ways. Most often, it refers to ‘Internet dependency’ or ‘digital dependency’ of individual people who appear addicted (unable to function emotionally) without being connected to the world wide web or some form of computerized device.

Beyond the psychological level, the subject of information age dependency has been largely ignored by social scientists, including in management studies involving cyber security, other than to refer to it as a vulnerability and then move on to a discussion of threats. In English at least, the scholarly study of information age dependency has been confined largely to the United States, and in that case it has been centred on a few locations: Idaho National Laboratory, Argonne National Laboratory, Sandia National Laboratory, and, to a lesser extent, Carnegie Mellon University. This is a classic case of a phenomenon long visible in universities where social and political reality, especially around new technologies, races ahead of the institutional capability of social science researchers in those or other universities to respond.

At the same time, in terms of the response of policy communities globally, the United States seems to be relatively alone in its focus on the subject. The Department of Homeland Security (DHS) has included the management of external dependencies as item eight in its ten points for planning for a ‘cyber resilience review’ (CRR) since 2009 (US Department of Homeland Security 2014). By 2016, Carnegie Mellon University had produced a resources guide on this subject ‘to help organizations implement practices identified as considerations for improvement’ during a CRR. This and the work of the three national laboratories comprise the bulk of research and analysis on the topic of cyber dependency.

This is a remarkable gap in global research and analysis even though the concept of dependency is a foundational premise for all security in the information age. In a literature review for the Australian Centre for Cyber Security, Thakur (2016) found that most of the research work had been done in respect of critical infrastructure resilience. Even then, ‘in contrast to the gravity of the vulnerability, the research reviewed reveals a lack of comprehensive information and analysis of cyber dependency that might clearly define the implications of it’ even in the case of critical infrastructure resilience. Thakur noted some ‘comprehensive modelling tool applications’, but saw these as ‘mainly theoretical or limited to closed environments’. She argued that there was an ‘absence of practical scenarios and multi-infrastructure analyses’ and that the challenge for researchers was novel: ‘scholars will need live and extensive cooperation from the operators of infrastructure and essential services, both nationally and globally, to reach the next level of comprehensive modelling and simulations that can provide insights into cyber dependencies’.
At the international level, a study on US/China relations in cyber space finds that there has been little comprehension in either country of the impact of cyber dependency on their broader security interests (Austin and Gady 2012). It gives a brief overview of the level of dependence, but this was captured succinctly and authoritatively in a July 2016 statement by China’s President Xi Jinping when he said that ‘our country is under others’ control in core technologies of key fields’ (Xinhua 2016). This was a staggering admission given that it had been Chinese policy for decades to work toward a self-sufficient high-tech industry, including in ICTs, and that China remained, in Xi’s own words, with weak foundations in science and technology.

We can ask, as the Austin and Gady (2012) report does, whether China’s level of dependence on a globalized ICT industry and international digital communications platforms is so high that it is forced to pursue cooperative behaviour on global cyber space issues rather than put at risk its international economic ties? The report also asks whether China’s economy can remain unaffected by the cascading effects of an extreme cyber attack on US economic targets of the sort that Americans fear that China might undertake. It concludes that ‘China is most likely obliged to cooperate in cyberspace rather than risk the fabric of its economic ties’ and that ‘China’s economy is almost certainly not immune from serious damage that could be brought on by a US cyber attack’.

More importantly, the report concludes that such questions could not be answered with a high degree of certainty because there were then few studies on the subject of shared cyber dependencies across international borders. That remains the case today with few studies documenting such dependency in any detail, and even fewer analysing its transformative potential on geopolitics or economic power.

There is wide recognition of the general condition of dependency. For example, a RAND Corporation report recounts discussions between Chinese and Americans on the issue of China’s dependence on the United States in the Track 2 talks sponsored by the Center for Strategic and International Studies beginning in December 2009 (Harold et al. 2016). The RAND report also concludes that China may have more to lose in economic terms than the United States if cyber disputes are allowed to escalate: ‘Chinese representatives cited numerous ways in which their country was dependent on US capabilities’.

This new form of mutual dependence, the mingling of interests and activities in cyber space affairs, is so profound that it has been called ‘entanglement’ and, in broad terms, this characteristic is shared among all countries. It is not possible to know how transformational the Internet may have been on geopolitics because there are no clear data on just how intermingled the critical elements of economic life have become.

The military security dimension of this dependence is a novel challenge. In the case of the United States and China, elements of the civil infrastructure almost totally reliant on the cyber domain (mobile communications, the Internet, electricity grids, landlines, undersea cables, banking) are also inter-mingled with military assets. The military power of states is now becoming more cyber dependent than ever, though the depth and implications of this dependence vary widely from country to country.

One of the best studies of international cyber dependence that begins to approach the granular level necessary is a 2010 report on the Reliability of Global Undersea Communications Cable Infrastructure (ROGUCCI) that was supported by several international stakeholders, such as the Institute of Electrical and Electronic Engineers, cable repair companies, cable-owning companies, and financial market representatives (Rauscher...
2010). Published by the EastWest Institute in support of the goal of promoting greater cooperation on cyber space security, the report aimed to document the reliability of the global cable network, assess the potential implications of failures in it (with a focus on the financial services sector), and make recommendations for enhanced security and resilience.

The ROGUCCI report is especially noteworthy because it points to the cross-domain character of cyber dependencies: engineering and maintenance problems can carry consequences for international relationships in business and politics, and even in personal relationships. Delays in diplomatic approval for foreign cable repair ships to enter the territorial sea of a host country to repair an undersea cable can have serious repercussions for the transactions relying on that access. The report notes that the overwhelming share, possibly as much as 85% of all Internet traffic, passes through undersea cables. While some localities are blessed with multiple undersea or land-based cables carrying Internet traffic, many are not, and there are worrying concentrations (single points of failure) in the global network around the northern Arabian Sea and in the Luzon Strait.19

Several years prior to the cable study, in 2007, the US government became concerned about such international dependencies in a range of critical infrastructures, including ICT-related, and set up the Critical Foreign Dependencies Initiative (CFDI). As part of this process, it asked its embassies to report on those facilities in their host country that might be considered critical for the national security or economic prosperity of the United States. Based on the responses, the DHS compiled a prioritized list of these dependencies, which included ‘over 300 assets and systems in more than 50 countries’ (US Department of Homeland Security 2014). In many cases, the landfall stations of the undersea cables in foreign countries were included, as was other telecommunications infrastructure. But at that time and since, as Arce (2015) points out, there has been little consistent analysis of the character of the vulnerabilities and risks (what I would call the ‘dependencies’). He also notes an ‘absence of cyber entities in the list’.20 The main enduring significance of this list might be that it highlighted the ‘discontiguous and non-traditional character’ of US vulnerabilities.

In 2014, in a report on global financial market infrastructures, the Bank for International Settlements described the management of dependencies (‘complexities and interdependencies’) as the biggest security vulnerability in the cyber domain (Bank for International Settlements 2014). By October 2016, the Group of Seven (G7; 2016) of the most developed economies was emphasizing the need for greater attention to intersectoral dependency involving financial services. It called on jurisdictions to ‘identify … interconnections, dependencies, and third parties … prioritise their relative importance, and assess their respective cyber risks’.

If indeed the cyber age is transformational of nearly everything, and if it is a multi-level (cross-domain) problem involving states, international organizations, private corporations, citizens, and communities, and if the study of one of its most fundamental aspects (dependency) is at an early stage, then we have some way to travel in understanding (rather than simply observing) the social construction of security in cyber space. The pleasant surprise the world experienced when social media played such a powerful role in the Arab Spring can now be contrasted with the dismay and uncertainty many people feel—after the apparent deceit surrounding Brexit and Donald Trump’s ‘alternative facts’—about the manipulation of cyber space and its information platforms in bringing into question the very legitimacy of democratic political processes in Western liberal democracies. The depth of surprise and
dismay about the negative impacts on politics of the information age is directly attributable to the lack of prior granular study of the dependency of power holders and their challengers upon the relevant types of cyber age phenomena.

**Cyber Dependency: A Deeper Look**

As dependency studies have improved, there has been a broadening of the lens as to what is involved. Table 2.1 compares the headline categories for the different types of dependency, in the specific case of critical infrastructure, from two different studies. Of special note, the cyber dependency is seen as sitting alongside other non-cyber dependencies, almost in a mutually exclusive position, without significant reflection on how the cyber age might transform the mitigation potential (or vulnerability potential) associated with non-cyber types of dependency.

There has been substantial development of the concept of dependency based in part on this early work. In an ideal world, assuming perfect knowledge, the end point of study of any type of dependency might be similar to that proposed by Petit et al. (2015) in respect of critical infrastructure. This goal is a comprehensive ‘understanding of all dependency and interdependency dimensions’ in a way that ‘allows decision makers to anticipate and characterize, in real time, how all dependency and interdependency dimensions influence the resilience and protection of a critical infrastructure system, of a region, and, ultimately, of the Nation’ (p. 24).

On the one hand, it is easy to agree with the view of a European Union directive that concluded that many of the dependencies are difficult to analyse in any detail: ‘participation in the global infrastructure ecosystem is inherently predicated on acceptance of a measure of unknowable risk’. On the other hand, as mentioned earlier, Carnegie Mellon has developed a manual (‘resource guide’) on external dependency management for cyber space. While

| Table 2.1 Headline descriptions of classes of dependence for critical infrastructure |
|-----------------------------------|---------------------------------|
| Rinaldi et al. \( ^a \)                  | Pederson et al. \( ^b \)            |
| Physical                              | Physical                        |
| Cyber                                 | Informational                   |
| Geographic                            | Geospatial                      |
| Logical (seen as human decisions or actions) | Policy/procedural          |
|                                        | Societal                        |


recognizing difficulties in data acquisition, the guide assumes that it is possible to identify, prioritize, monitor, and track external dependencies, even if 'organisations have a limited ability to directly monitor and control the vulnerabilities and threats introduced'.

It is of considerable philosophical and ontological import that the analytical processes themselves, intended to enhance security and resilience in cyber space, demand (if only for the purposes of dependency analysis on a continuing and timely basis), the creation of entirely new sets and types of social relations, at some or all levels of social organization (individual, community, national, and international). For example, at the international level, in the interests of promoting cyber security, leading intergovernmental groups and international organizations, such as the BIS mentioned earlier, have called on states to share information on their dependencies, but this has involved a process of trust building and new relationship creation, both of which have become even more difficult under the pressure of escalating international tensions over cyber intrusions or abuse.

Understanding dependency is not just an information problem. It is a knowledge problem and furthermore a knowledge management problem, and both these sets of challenges also introduce new social realities and constructs. On the first point, the mapping of dependencies and the study of the linkages require the 'combination of multiple areas of expertise'. The need to create, marshal, and exploit these diverse levels of expertise is itself a new phenomenon that creates new social structures as well as new opportunities and tensions between communities and states.

The second point about knowledge management of dependency is profoundly more interesting and challenging. How can a single political leader or corporate manager visualize any dependency that is constituted by phenomena that are highly divergent in character, plane of action, and speed of manifestation. Many cyber dependencies are simultaneously constituted, on the one hand, by a global geographic scope in terms of Internet-based systems that they use (and the physical human activities set in train by the systems) and, on the other hand, by invisible, atomic level electronic pathways and data that can produce system failures in milliseconds under multi-vector, multi-phase attacks, or malfunctions. According to specialists from the Argonne National Laboratory in the United States, writing in 2015, no scalable approach, standardized capability, or combination of capabilities to undertake such a mapping currently exists. The authors propose 'a critical infrastructure dependency and interdependency assessment framework' that can evolve over time, be based on flexibility in approach, and 'allow the implementation of innovative capabilities that will reflect the evolution of technical capabilities and of critical infrastructure protection and resilience policies'. Even after the processes to collect the relevant data and analyse them are executed, the process of synthesizing and communicating them presents a novel challenge. Just one part of this challenge is the question of visualization of the map of dependencies in a way that is meaningful to decision makers who have had no part in the data collection or analysis, bearing in mind that the visualization is operating across a potentially large geographic scale at the same time as needing to incorporate assessments of non-tangible electronic activity. The authors say that 'Development of GIS visualization capabilities is vital for the analysis of critical infrastructure dependencies and interdependencies, especially in visualizing cascading and escalating failures at the regional level' (beyond a small locality involving one or two assets). They foreshadow a capability that can integrate results for generating forecasts of pathways of cascading,
escalating, and common-cause failures in a way that can address second- and third-order dependencies.\textsuperscript{25}

Thus, the process of addressing dependency in cyber space creates new information products with hitherto unseen levels of sophistication and complexity. Arguably, the need for new products designed to map and understand highly entangled dependency may actually aggravate it, thereby frustrating the very logic and purpose of such dependency study.

The proposition that the concept of cyber dependency analysis is in its infancy, as suggested by Petit et al. has been borne out by initial results from the roll-out of a campaign by the DHS to get the country’s corporations attuned to the management of their external dependencies. We can source DHS work in this area to around 2009 and possibly earlier but, by January 2015, in reporting results of a joint DHS/Carnegie Mellon survey, Gaiser and Haller (2013) concluded, based on a sample of US enterprises, that the country’s maturity in this area of activity on a scale of 0–5 was probably below 1. And the United States is by far the most advanced country in understanding of cyber dependencies.

In general terms, security in cyber space remains very weak at best and in most places is non-existent. This applies as much to our dependency at the personal level on unverified sources of information and on emails as it does at the national level to management of critical infrastructure and preparations for cyber-enabled war. As Cornish observed, ‘complete dependence on the cybered world is generating a complacency and fatalism like that caused by the complete vulnerability to MAD’,\textsuperscript{26} a reference to the nuclear warfighting doctrine of mutual assured destruction.

**Conclusion: Towards Political Framing of the Infosphere**

For all the wonders of the information age, the dangers may for now be greater. This prospect arises from one simple consideration foreshadowed by Bell in 1976. We do not have a social and political design that can keep up with the pace and character of change in the information age. Security in cyber space has only recently become a serious national and international policy consideration, and smaller communities and individual citizens are just beginning to come to terms with the issues. This chapter argues that there is an inevitable linkage between these four levels of analysis. It also argues that the infosphere has created new social and political realities that are potentially transformative. Of these, issues of dependency are among the most important but also among the least studied. The primary challenge in policy terms is for institutional and social renovation that addresses the challenges of distributed political authority and the ethical and practical implications of the reaggregation of data. As Bell characterized it, we need more scientists (including social scientists) to compensate for the ‘talented tinkerers’ in technology. The social science of security in cyber space is only at an embryonic stage of development and is not keeping pace with the extraordinary work of both the gifted scientists and the talented tinkerers.
Notes

4. Castells, Networks, 239.
7. Floridi, Ethics of Information, 296.
8. Personal communication with the author.
10. Floridi, Ethics of Information, 275. As I read it, Floridi is not postulating what may be right or wrong in particular cases but rather that in the information age a judgement on the moral worth of an enabling act depends on how well it conforms to the inherent moral character of the information process, which is its power to create a desirable kind of distributed morality.
11. Floridi, Ethics of Information, 274.

References


Digital technologies are so prevalent now, and so intertwined with many aspects of our lives, that it is easy to forget just how rapidly they have emerged and how profoundly they have had an impact upon us. Technological transformations always have implications for the way people work, play, live, and die. The first and second industrial revolutions (mechanization of the textile industry in the late 1700s and steel/electricity in the latter half of the 1800s) brought about the complete reorganization of societies from self-sufficient, largely agrarian communities into urbanized, interdependent communities. Unlike previous technological shifts that were absorbed slowly over many decades or even centuries, these periods of intense change had significant impact upon people's lives within a single generation.

Looking back, of course we can identify a whole range of political decisions, pressures, expectations, and power that facilitated and promoted the industrial revolutions—quite distinct from science or technology. In large part, these political levers also shaped the way those transitions emerged and developed. We now find ourselves living through the third Industrial Revolution: the Information Age; and, arguably, embarking upon the fourth: the Internet of Things. Benefitting from historical examples and equipped with the understanding that politics not only shapes the way technology develops but also that technology itself may have significant implications for how political practices are carried out, we can take a more analytical and critical approach to this latest transformation. This chapter highlights some of the ways that politics and digital technologies have affected one another over the past quarter-century—specifically in the context of cyber security. In doing so, it argues that power shifts in what may seem to be largely technical domains often have quite significant correlations in global politics. For this reason, they are worth observing and analysing with some care.
There are a number of ways that an awareness of the political history of cyberspace can help us understand and interpret technological change and its implications for society. First, it can help us to understand why certain technological decisions were taken—for example, the adoption of one standard over another. This is particularly interesting in terms of cyber security where we have observed the deliberate adoption of some protocols, practices, and standards that were recognized—even at the time—as difficult to secure (Willemssen 2001). Sometimes, these decisions have been taken wholly within the technical community, with no obvious link to political preferences. In other cases, these decisions have actually been the subject of intense political debate. And in yet other cases, while politics may not have been a direct or overt influence, we might observe some underlying tensions or preferences that we can trace back to political ideals and values.

The second way that an awareness of the political history of cyberspace can help us understand and interpret technological change and its implications for society is by providing background to contemporary political debates. Negotiations in the United Nations over cyber norms, proposals to limit state aggression in cyberspace, and talks between heads of states about cooperation in fighting cyber threats all take place today on the back of considerable dispute, contestation, and also some cooperation at a political level. Without understanding something of the political history, these negotiations can be misinterpreted or appear disconnected from the context that has informed them.

Finally, the political history of cyberspace can be useful in articulating exactly how political actors can shape the future of technology. In most histories of cyberspace or the Internet, the focus is on technological milestones or developments. This is important but it often serves to mask the motivations and intentions of political actors who seek to influence technological developments and adoption. All these aspects of the complex intersections between technological developments and global politics are essential to considerations of how power might be redistributed in the future, and what that might mean for global peace and security.

Chapter 4 further develops these points through a series of steps. First, some of the key political events that ultimately shaped and promoted the development of the Internet are outlined. These help to illustrate the ways that technological developments are linked to, and sometimes embedded within, political tides and currents. The second section runs through some of the evolutions of political perceptions of cyber security threats. Observing both change and continuity in these perceptions helps us put current approaches into some perspective, and avoids the two reductionist arguments that either ‘everything is different now’ or ‘it is the same as it ever was’. The chapter then turns to some of the ongoing debates about how to conceptualize this technology—particularly with relevance to the political implications it evokes. It is necessary to have some understanding of these in order to engage with the final section, which looks at which remedies have been proposed thus far—including those that have had some success as well as those that have not moved forward. Essentially, the chapter calls for much more careful and comprehensive engagement with the interrelationship between technological developments and political forces. Without a better understanding of these dynamics, it will be unlikely that we will move through this technological shift with any better prospect of delivering positive, human-centred outcomes for society than was the case in the first two industrial revolutions.
The Political Evolution of Internet Technology

This edited volume very usefully dedicates several chapters to the much-needed examination of perspectives on cyber security from non-Western countries. This has been a real deficit in the past few decades of scholarship and one that it is essential to address if we are to better understand the complexities of global cyber security. Cyberspace has such broad implications and intersects with the human condition, politics, the economy, military power, and other global affairs in such important ways that it is not at all surprising that not all communities regard its benefits and challenges in the same way. Some political cultures place more emphasis on state security and others on personal security. Some countries are deeply reliant upon cyberspace while others are much less so. And, crucially, it is not at all clear that state power in the Information Age can be as clearly recognized and exercised as it was in the Industrial Age, which means that our ideas about which states are powerful and which are vulnerable may need reconsideration (Carr 2016a). In order to come to grips with the implications of cyberspace for questions about power, equity, legitimacy, and vulnerability in the future of global politics, we must engage much more comprehensively and honestly with non-Western views than we have in the past three decades.

This chapter, however, dealing as it does fairly narrowly with the political history of cyberspace, focuses largely on the United States. While there was certainly plenty of activity of both a political and a technical nature taking place in other states, throughout the latter half of the past century, the bulk of activity in both these domains was definitely in the United States. This was in part a consequence of heavy investment by the US government, excellent research institutions, and also the political vision of some key actors. There can be little question that, in the early second half of the twentieth century, the United States displayed considerable leadership on Internet technology. Crucially, it also dedicated concentrated attention to the integration of emerging technologies with a vision for how best to strengthen and project US power. To understand the political history of cyberspace, it is therefore necessary to engage in some depth with the political history of cyberspace in the United States. The critical lesson here is that, when states successfully link their strategic ambitions to technological investment, this can profoundly affect their global power status. This is a point worth keeping in mind as one reads through the subsequent chapters in this volume—particularly those that delve into non-Western states.

Catalyst: Sputnik Crisis

There had been a long-standing and strongly held belief in the US that technology and power were connected. Manufacturing technology had transformed the fortunes of the US through the first and second industrial revolutions. Transport and communications technologies had been essential to colonizing the west and uniting the expansive landmass into one ‘United States’. In addition, military technology had allowed the United States to establish its dominance over the Western hemisphere through the Monroe Doctrine. This belief in
technology as linked to state power meant that confidence in US global leadership was severely undermined by one event in the late fifties.

On 4 October 1957, the USSR successfully launched the Sputnik satellite into orbit. The Sputnik success was a profound shock to the US, which also had the (less ambitious) Vanguard satellite programme underway but still a long way from maturity (Killian 1977). It was clear that the USSR had surpassed the US in a critical technological moment which, in the context of the Cold War, had very serious implications. Satellite technology was expected to be married eventually to missile technology—potentially to allow for nuclear strikes from space (Dickson 2001). The United States government responded by investing heavily in the establishment of both the National Aeronautics and Space Administration (NASA) and the Defence Advanced Research Projects Administration (DARPA; Abbate 2000). DARPA (or ‘ARPA’ as it was renamed) would become a central source of funding for academics and industry researchers focused on networking and related technologies. The very proactive support channelled through ARPA was fundamental to the United States emerging as a global leader in information technology (IT)—a point often overlooked in assertions more common today that technological innovation is the domain of the private sector.

The Cold War Research

Throughout the remainder of the Cold War years, the United States continued to invest enthusiastically in technology. While there was no way that the transformative effects of the Information Age could be anticipated back then, the computational power to solve complex problems and process data was seen as essential to most areas of science and technology that might underpin a return to global technological dominance for the United States. For this reason, harnessing this computational power for as many projects as possible was regarded as efficient resource management, and finding a way to get the most out of expensive computers that were typically housed in universities or large research institutes became a priority. Networking these computers would allow better access and support more work, and it was funded generously by ARPA on that basis.

It is important to note that, while the US government saw the links between strategic competition with the USSR and investment in technology that could be channelled to developing a military advantage, the researchers funded to carry out this work were certainly not all motivated along those lines. They were largely mathematicians, physicists, programmers, and engineers who developed research projects along the lines of their own interests while making use of available funding.

One of the projects to emerge from this research funding was a proposal to develop a secure and resilient communications system that was envisaged to enhance the second strike capability of the United States in case of a nuclear attack. The existing telephone system was a ‘hub and spokes’ model, which meant that, if the ‘hub’ (the exchange) was disabled, the system as a whole would be incapacitated. It seems a prosaic concern today, but the reality was that, in the event of a nuclear strike on the United States, a phone call would be necessary to launch a retaliatory attack and thus the ‘second strike’ capability rested upon a
vulnerable communications system.¹ At the height of the Cold War, this was certainly a valid concern and, thus, the project was funded. It contributed to the development of networking technology, in concert with the many other funded research projects undertaken through those years.

Although computer scientists around the world were working on developing networking protocols at the same time, in the United States, this government funding was fostering a core group of talented researchers including Robert Kahn and Vinton Cerf who would later be credited with developing the important Transmission Control Protocol/Internet Protocol (TCP/IP; Cerf and Kahn 1974). This concentrated post-Sputnik investment in US technology research had exactly the desired effect: there emerged a concentration of skills and knowledge that gave the United States an edge and allowed it to shape the future of computer networking to a large extent.

**Clinton/Gore Initiatives and Vision**

By the beginning of the 1990s, a number of important changes had taken place that were significant for the political history of cyberspace. The Cold War had concluded with the dissolution of the USSR and a new team of Democrats had moved into the White House. Bill Clinton and his vice-president Al Gore formed a cohesive and effective political partnership. Clinton's ambitions to turn the world beyond the Iron Curtain into a truly global marketplace for the United States married neatly with Gore's long-standing belief in the future of IT (Gore 1989). Together, they formulated a powerful plan to use the internet to deliver US goods and services to the world, and to spread democracy and human rights—both in an effort to reshape the post-Cold War global order (Clinton 1993).

As was the general trend in many Western states in the 1980s and 1990s, the Clinton/Gore administration had a strong preference for private sector ownership of critical infrastructure—of which information and communications infrastructure was one component (Carr 2016b; Legrand 2014). Although the internet had been government funded and directly or indirectly managed by the state from its inception, the Clinton/Gore vision was very much one in which both the infrastructure and the management (what came to be known as ‘governance’) of the Internet would eventually be taken over by the private sector. However, it was not entirely clear in those early days just what the business case for the provision of Internet access might be and the private sector was cautious about committing investment. It was only after years of government investment, management, and strong political belief in the future of the internet that private organizations began to see that there was, indeed, commercial benefit to be harnessed. In keeping with the Clinton/Gore vision, by the mid-1990s, the internet was both commercialized and privatized.

These dominant political ideas in the United States would further shape the way this technology developed with lasting consequences for cyber security. The preference for openness, interoperability and innovation without permission were arguably fundamental to the success of the internet—and they also helped to build in some of the intractable security problems with which we now contend.
The Evolution of Cyber Threats

As information and communications technology developed and as states’ reliance on this technology increased in scope and sophistication, the imagination and innovation of malicious actors in cyberspace meant that threats and attack vectors continued to evolve at a relentless pace. Consequently, ideas about what constituted cyber security vulnerabilities, and who the primary threat actors were, also changed. Looking back at the ways that cyber security has been perceived by politicians over the past 30 years provides some interesting insights into how the field has developed and evolved. It also provides some perspective on change and continuity that might help us anticipate future trends.

Early cyber threat actors were typically computer enthusiasts or engineers—very often motivated by proving either that they had the skills to break into a system believed to be secure (a kind of self-aggrandizement) or that the security flaws that they were able to exploit should be taken more seriously (a ‘public good’ motivation). This activity often took the form of intrusion and the proliferation of malware like the Morris worm or the Love virus was indicative. These were regarded as serious and concerning but they were predominantly viewed through a law enforcement lens and not a political or global conflict lens. Both these activities obviously continue today on a vast scale, but they are now augmented by much more sophisticated practices that move beyond criminality and into the domain of political conflict.

Early website defacement

In the very early years of thinking about cyber security as a national security concern, there was a focus on website defacement. This tended to take place in the context of existing political tensions. Pakistan and India, or Palestine and Israel, were common cases of website defacement. Although it seems relatively inconsequential today, politically motivated attacks on websites were initially seen as a new domain of political conflict. There was a kind of anxiety that stemmed from the realization that these attacks were difficult to trace back to a particular person, that they could arise from beyond the state borders, and that they could be used to inflame political conflict in the physical world or politically influence civil society.

Cyber Pearl Harbour

These same anxieties featured in the next wave of perceived cyber threats: distributed denial of service (DDoS) attacks on critical infrastructure. Critical infrastructure vulnerabilities had been a concern for policy makers almost from the beginning of the internet. As critical infrastructure like utilities, transport, and communication were denationalized and sold off to private sector owners and operators, and as networking technology developed, the inevitable happened. The industrial control systems used to control critical infrastructure were networked. This allowed for remote monitoring and management, and was seen as one way to enhance efficiency and cut costs. However, for various reasons, these
systems were relatively vulnerable to attack. They tended to be built on outdated and insecure operating systems, and it could be both costly and risky to install software patches and updates—a fundamental element of contemporary cyber security practices (the May 2017 ‘Wannacry’ NHS incident is a key example of this). This led to the use of the term ‘Cyber Pearl Harbour’—particularly in the United States—to depict the political anxiety about a potentially devastating surprise attack on the homeland. The perpetrators of such an attack were expected to be a mix of actors: determined non-state criminally or politically motivated groups (including terrorists) or possibly states.

In 2007, in the midst of a diplomatic row with Russia over the relocation of a military memorial in Tallinn, Estonia experienced just such an attack, causing disruption to governmental websites and many services like news dissemination, banking, and law enforcement. Although this DDoS attack has never been conclusively attributed to the Russian state, the political context that enveloped it led many to believe that this was indeed a state-based attack. Estonia’s appeal to NATO for support brought into focus the alliance’s inability to deal with this new security threat under Article Five of its founding treaty and it prompted a wave of new thinking about how cyber insecurity could be understood within conventional international relations and international law frameworks.

The experience gained by both Estonia and NATO prompted governments, alliances, and regional groupings to consider much more carefully how exactly they might respond, cooperate, and work to mitigate cyber attacks in the future. The development of cyber security doctrine and policies intensified in many states and NATO established the Cooperative Cyber Defence Centre of Excellence (CCD CoE), which has subsequently produced some leading research on these issues, including the Tallinn Manual on the International Law Applicable to Cyber Warfare (Schmitt 2013).

Stuxnet

In 2010, another significant cyber incident took place that further shaped perceptions of the politics of global cyber security, especially in relation to critical infrastructure. Iran reported that it had suffered a targeted cyber attack on the industrial control system at its Natanz nuclear enrichment facility. The Stuxnet ‘worm’, as it came to be known, facilitated the physical damage of a set of centrifuges at the facility, reportedly disrupting the Iranian nuclear programme. This incident was significant because it was the first publicly reported example of a cyber attack causing physical damage to what could be regarded as a military asset. Once again, technical attribution has not been possible (at least not within the public domain) but, as in the Estonian case, the political context led many to conclude that the perpetrators were identifiable. In this case, the attack was believed to come from a joint programme between the United States and Israel—an account that has been comprehensively documented by New York Times journalist David Sanger but not officially verified by either state.

Internal challenges

While attention had been somewhat diverted to state-based activity as the source of threats in cyberspace, a new internal security concern emerged in the United States that
had implications for global politics. In 2010, the whistle-blower website WikiLeaks made available a large number of US classified documents to major media outlets—documents mainly relating to the Iraq war that had been exfiltrated from a military network by Bradley (later Chelsea) Manning—then a US soldier serving in Iraq. Three years later, the Central Intelligence Agency IT contractor Edward Snowden also released classified details—this time about the extent of the US intelligence community’s surveillance capability and practices. Not surprisingly, there are quite polarized views about whether these acts were commendable or condemnable. Although there are some important distinctions between the two cases, both of them reinforced the growing sense that many people are conflicted about issues such as information transparency, secrecy, and privacy—both at an individual and a state level. In the context of the Information Age when so much data can be stored on such physically small devices, it appears to be increasingly difficult for states to conceal much. Choosing what to conceal and what to reveal has been a fundamental element of statecraft and, thus, these challenges of controlling information, data, and secrets are being regarded as cyber security concerns in global politics to a new and enhanced degree.

The Future of Cyber Security Threats

Of course, the cyber security concerns of yesterday and today might not necessarily be the concerns of tomorrow. This brief summary of changing threat perceptions over the past few decades makes that abundantly clear. We have some insight into what will possibly become a further cause for concern, but predicting the trajectory of technological change and especially predicting the way that humans will interact with changing technology has always been more of an art than a science. There are many anticipated vulnerabilities that researchers, analysts, and policy makers are already focusing on, and others will emerge over time. Some of the issues that look set to feature in the future of international security are to be found in the developments of what is called the ‘Internet of Things (IoT)’ or the ‘fourth industrial revolution’.

The Internet of Things refers to the increasing proliferation of embedded sensors in devices, in our environment, even in our bodies—that collect and transmit data that is then analysed and often used to produce real-world effects. There are expectations of significant benefits in the collection of the vast data sets that will be possible as a consequence of networking the hundreds of billions of devices that are expected to comprise the Internet of Things by 2025 (Statista, 2021). These benefits include not only economic value but also considerable research potential. For example, the collection of data from implantable medical devices can provide a scale of valuable research material otherwise impossible to collect through conventional means. However, there are many questions about the rights, liabilities, and responsibilities of personal data. Who will own the data generated by implantable devices? The person in whose body it resides? The medical practitioner who implants and monitors it? The company that manufactures and maintains the devices? And to what extent can people give consent to the use of their data when, in the Internet of Things, one is not always aware that one is sharing data or even interacting with a device?

Intrinsic to the cyber security concerns evolving around the Internet of Things are the ‘real-world effects’. Devices that are connected can be compromised. Security researchers have already demonstrated how cars can be hacked and intentionally crashed. The same
security implications apply to implanted or ingested medical devices and anticipated autonomous transport systems like international shipping. Also intrinsic to these evolving cyber security concerns is the growing and ever more critical reliance on data integrity and data privacy. If systems at both the micro (our bodies) and macro (critical infrastructure) levels increasingly rely on the automated collection and analysis of data, it becomes ever more important that we can rely upon the integrity and availability of those data streams. Another interesting dynamic here is the renewed challenge of DDoS attacks, briefly regarded as of little real concern due to the fact that they are not sustainable over a long period of time and do not usually cause lasting damage or facilitate theft. However, every insecure and connected device is an entry point to the network through which malicious actors can access and disrupt legitimate activity. Given that many Internet of Things devices are so small and so simple that they cannot support security mechanisms like passwords or software updates, there is some concern that these hundreds of billions of (largely unsecured) devices are rapidly causing the decay of whatever network security we have managed to achieve to date. And once these devices are coordinated into a DDoS attack (we witnessed the first one in late 2016\(^2\)), the potential for widespread disruption escalates quickly.

It is not possible (or necessary) to describe every past or current cyber security vulnerability here. Nor is it possible to provide much certainty about the future of cyber security—although many chapters in this volume will offer considerable insight into that very question. In seeking to anticipate future security vulnerabilities and to mitigate against them, an understanding of the historical trends and political perceptions of cyber security can help provide an awareness of both change and continuity. We can see that some threats relate to concerns about the resilience of the state, some relate to civil society concerns about privacy and security, and others combine both.

**Major (and Ongoing) Debates About Conceptualizing Cyber Security**

Although there has been general agreement that the problems and opportunities of cyberspace require coordination and cooperation across sectors, across borders, and across levels of governance, putting this into practice has been quite difficult. There are many reasons why, including the problems of coordinating domestic law and applying international law, the difficulties of conclusively attributing malicious activity in cyberspace, and the wide range of (sometimes competing or conflicting) interests that need to be accommodated. Also relevant, and perhaps less discussed, are quite fundamental differences in how key principles or concepts are understood and internalized by different state actors. These differences can quite significantly shape approaches to cyber security and the differences discussed later have been key factors in attempts to fashion a more coherent global approach to cyber security. It is important to note, however, that, although these factors are presented as dichotomies here, this can also dangerously over-simplify complex positions and considerations that further shape the approaches of state actors as well as the interactions between them. In reality, most political actors are not situated neatly at either end of these polarized positions but, rather, somewhere in the grey area between them.
Cyberspace as Borderless or Sovereign?

Internet technology heightened and intensified discussions that emerged from the globalization literature about the utility and ongoing relevance of the Westphalian system of states (Eriksson and Giacomello 2006). This (non)alignment of the territorial state with the internet is one of the divisive issues that underpin a lot of debate about global cyber security. There are two powerful reasons why some argue for the promotion and protection of a universal online experience for all those who access the internet. First, it has come to be regarded as a human right that all people should have access to the most comprehensive knowledge base ever conceived. If only some had this privilege, they would surely benefit disproportionately to those who did not—thereby entrenching the ‘digital divide’. The second reason is that it is believed by some that the very act of connecting civil society through the internet could be a mechanism by which to promote peace and unity—through a newly connected ‘global civil society’ that would be able to circumvent or supplement conventional great power politics for the betterment of humankind. These ideas have been very firmly grounded in the United States, the European Union, Canada, Australia, and other ‘like-minded’ states which tend to refer to cyberspace as a ‘global commons’ or ‘global public good’. Indeed, these ideas were eventually folded into US foreign policy under Secretary of State Hillary Clinton who combined them in the US ‘Internet Freedom’ doctrine (Clinton 2010) and, in 2012, Internet Freedom was declared a universal human right by the United Nations (UN Human Rights Council 2012).

For many other states, the perceived necessity of controlling both the content available online and activity carried out online, has led to a reassertion of sovereign principles of state authority. These states regard cyberspace as very much an extension of territorial space and, therefore, subject to the same levels and dimensions of state control as physical space. Of the great powers, sovereignty in cyberspace has been promoted most assertively by China and Russia but it also resonates strongly with many others—particularly post-colonial states. For these states, any relinquishment of sovereign control can be regarded as an unwelcome throwback to pre-independence. Contrary to some suggestions that this is simply the view of oppressive or autocratic regimes, it can be quite a complex issue that eludes a binary narrative of ‘free’ or ‘not free’ (Cornish 2015). In fact, civil society in these countries can often be supportive of firmer state control because they regard economic growth and national autonomy as indicators of governmental legitimacy (Mueller and Wagner 2014).

Information Security Versus Cyber Security

Related to the conceptions of cyberspace as borderless or sovereign is a second conceptual polarization: that of information security versus cyber security. At the great power level, these two concepts again tend to define the Chinese/Russian and the US approaches respectively. The United States prefers the term ‘cyber security’ to indicate a more confined mandate limited to infrastructure and applications. ‘Information security’, which Russia and China use, encompasses content as well—particularly content that could be politically or socially destabilizing. For those states, controlling politically or socially charged information is seen as essential to maintaining internal cohesion and it is therefore linked to national
security. States that favour ‘cyber security’ as the defining concept argue that the ‘information security’ approach can be a mechanism for political oppression and is antithetical to the principles of a free, open internet that they favour. The problem of defining and enforcing jurisdiction in cyberspace remains deeply problematic. Questions about who can or should control information continue to be debated, contested, and reconsidered as the implications of cyberspace shift and change.

These debates, of course, link back to the tensions over cyberspace as borderless or sovereign but they also relate to questions about the public/private layers of responsibility and authority in controlling internet infrastructure and content. The complexity of these arguments played out in a very stark and significant way during the 2016 US elections when there were allegations that Russia intervened to shape the election outcome. This suggestion that one state could potentially manipulate information in another state to pursue its own interests is one that has been raised many times at the United Nations—initially by Russia in 1998 and repeatedly over subsequent years by both the Russians and the Chinese (Tikk-Ringas 2016).

Security versus Privacy

Unlike either of the above conceptual tensions, different understandings of security and privacy are not as easily ascribed to particular states. Rather, they involve complex, intertwined and sometimes conflicting ideas that are distinct to actors, sectors, and contexts. Both security and privacy are abstract and deeply subjective concepts. For some actors, and in some contexts, privacy and security are indivisible. Surveillance and the self (or state) censorship that it evokes are fundamental threats to security. If individuals cannot enjoy a private life, free from scrutiny, there can be further negative implications for the political health of a state. Civil society’s capacity to hold the government to account can be undermined and potentially lead not only to an individual sense of insecurity but a broader, national, and even international insecurity.

For others, online privacy or anonymity facilitates antisocial or even criminal behaviour. The challenges of connecting malicious activity to an individual actor can seriously undermine the efforts of law enforcement and can erode the sense of personal or professional accountability that, to some extent, moderates behaviour in a positive way. By this view, transparency is more important to security than privacy: the fact that the internet allows actors to shield their identity behind an avatar, a user name, or a re-routed IP address leads to a level of risk-taking and undesirable behaviour that can undermine the security of others.

Related to this are differing views about the appropriate role of the private sector in internet governance and cyber security. Again, many Western states that prefer minimal government involvement have regarded the private sector as better equipped to drive the direction of internet technology (though this is beginning to be challenged now). This is generally for several reasons. First, they tend to view the private sector as being in possession of superior technical expertise. Second, there is a prevailing view that private sector ambitions are somehow more legitimate and perhaps more impartial than those of governments. Others have significant concerns about, first, how the private sector actors use the information they collect and whom they sell it on to, and second, the extent to which they themselves cooperate with governments that wish to make use of that data. (For a more comprehensive overview of three different approaches to data governance, see Carr and Llanos, 2021.)
These complex approaches to security and privacy, and to the interplay or relationship between them, cannot be untangled here. Security studies (and critical security studies) and psychology are both rich fields that seek to understand these and other related concepts, and it would be impossible to devote anything but the most cursory attention to them here. However, an awareness of some of the nuances and divergent views of these concepts is essential to keep in mind when engaging with this material. Not only can these terms have different meanings for different actors in different contexts but, most problematically, they are often conflated and assumed to mean the same thing. ‘Cyber security’ and ‘information security’ are equally meaningless terms unless we first unpack the meaning behind ‘security’.

Proposed Remedies: Failed, Stalled and Ongoing

In considering the complex intersections between technological developments and global politics, it is useful to examine the ways that states have attempted to pursue a global, coordinated approach. While multilateral coordinated efforts (at least in the political domain) have been relatively slow to eventuate, what did develop quite quickly were domestic cyber security strategies. Here again, the United States led the way. Cyber security began to appear in the US National Security Strategy in 1998 under the Clinton presidency. In 2000, Clinton released the first National Plan for Information Systems Protection (Clinton 2000) and, in 2003, President Bush released The National Strategy to Secure Cyberspace (Bush 2003). Other states began to follow this lead, initially addressing cyber security through national security strategies and then, relatively quickly, producing dedicated cyber security strategies.

These national cyber security strategies provide insight into how governments have perceived and interpreted cyber security as an issue of national interest. They are useful for understanding how ideas about the politics of cyber attacks have changed over time, which big debates have developed and dominated the political discourse, and how future security concerns are taking shape. These are all central to the politics of cyber security. While these national policy documents continue to proliferate and provide insight into the commonalities and distinctions in state views on cybersecurity, a number of international initiatives have also been developed or proposed to try to address collectively the challenges of cybersecurity.

International Code of Conduct for Information Security

At quite an early stage, states began to anticipate the challenges that cyberspace might pose for global security. Concerns about the potential for ‘information weapons’ and the ‘threat of information wars’ were raised at the UN by Russia as early as 1998. However, the considerable challenges of integrating this rapidly evolving technology into existing concepts and practices of international relations mean that, despite decades of diplomatic effort, few resolutions on cyberspace and global security have been reached. In 2011, four states proposed for consideration by the UN a new International Code of Conduct for Information
Security for consideration by the UN. Russia, China, Tajikistan, and Uzbekistan argued that it was necessary to clarify the rights and responsibilities of states in protecting information infrastructure. The proposal placed a heavy emphasis on sovereign control and domestic legal jurisdiction. It also called for a multilateral mechanism to resolve international disputes in cyberspace.

The United States criticized the proposal for its emphasis on ‘information security’ rather than their preferred term of ‘cybersecurity’. In 2015, the proposal was revised to incorporate some of the feedback from the international community and reintroduced to the UN by an expanded group of states that now included Kazakhstan and Kyrgyzstan. Significantly, this group now constituted the full membership of the Shanghai Cooperative Organization. The revised proposal was unable to attract more widespread support though these debates continue through the UN Group of Governmental Experts process discussed in more detail elsewhere in this chapter.

Budapest Convention

In 2001, the Council of Europe introduced the Convention on Cybercrime (the Budapest Convention) and opened it for signature by all states. This remains the only international treaty on cyber security. It has had some success but has also encountered some persistent challenges. The aim of the treaty is to enable more effective cooperation between states (specifically, between law enforcement agencies within states) in investigating and prosecuting cyber crime, which is often initiated in one state and has effects in another. To do this, states must align their criminal codes so that there is a common agreement on what constitutes a crime in different jurisdictions. At the time of writing, the Budapest Convention has been ratified by 54 European and non-European states.

Some key states that have not signed or ratified the Budapest Convention include the BRICS (although, in 2001, South Africa signed but never ratified the treaty). Russia, China, Brazil, and India are among the states that have objected for two main reasons. First, for some states, the Budapest Convention raises concerns about the relinquishment of sovereign control. The Russians, in particular, have said they will not assent to permitting foreign law enforcement agencies to conduct internet searches inside their borders (Markoff and Kramer 2009). The second reason why some states have objected to the Budapest Convention has to do with procedural rather than substantial issues. The treaty was developed and implemented largely within the European context although there were a number of participant and observer states from outside Europe involved in the process. Some states are unwilling to sign up to a treaty that they had no voice in drafting. This need to balance inclusiveness with efficacy that can sometimes only come from working in a smaller group will also likely become an issue for the UN Group of Governmental Experts (UNGGE) that has been working in small groups to develop a set of proposed ‘cyber norms’.

United Nations Group of Governmental Experts

As noted in this chapter, Russia raised concerns about the militarization of cyberspace at the UN General Assembly in 1998 (Inkster 2015). In response to this, the United
Nations established the UNGGE to explore developments in the field of information and telecommunications in the context of international security under the Committee for Disarmament and International Security (First Committee). Its goal has been to agree some norms of responsible state behavior in cyberspace that could be universally accepted. The UNGGE comprises the five permanent members of the Security Council and a small group of other states that apply to take part. It convened for the first time in 2004–5 with 10 additional states but the group failed to produce a consensus report. It then reconvened in 2009–10 (10 additional states), 2012–13 (10 additional states), 2014–15 (15 additional states), and 2016–17 (20 additional states and still underway at time of writing). The 2014–15 UNGGE produced an important consensus report in which the participants identified 11 proposed norms for responsible state behavior in cyberspace. Among these are norms that prohibit certain behavior, including ‘states should not knowingly allow their territory to be used for internationally wrongful acts using ICTs and states should not conduct or knowingly support ICT activity that intentionally damages critical infrastructure’. In addition, there are norms that speak to best practice or expectations of actions that states should take. These include the normative claims that ‘states should take appropriate measures to protect their critical infrastructure’ and that ‘states should respond to appropriate requests for assistance by other states whose critical infrastructure is subject to malicious ICT acts’.

At this stage, these norms are only proposed norms and considerations on how best to implement them continue. One line of discussion involves more specificity about what exactly the 11 norms entail in practice. Another involves considering how they might be more widely adopted. Presumably, the 20 states involved in delivering the 2015 consensus report agree to adhere to those norms but they now need to be internalized by the rest of the international community. An update is due in 2021 that will develop this further.

The UNGGE norms process is arguably slow and is frequently criticized by external actors, especially in the private and technical sectors, for being too disconnected from the day-to-day realities of cyber security. However, it could be argued that the norms process is an essential step for international coordination and cooperation on cyber security, and that the consensus report and the 11 proposed norms are actually quite a significant achievement—given the range of complex factors that inhibit states aligning more closely over cyber security. Indeed, over the past 25 years, neither technical nor legal solutions have proven much more effective. Despite a hugely expensive and profitable cyber security technical industry, we currently find ourselves as insecure as, or possibly more insecure than, ever before. In terms of legal solutions, the application of international law to cyberspace has proven deeply problematic. At this point, and despite much effort, there is no agreement on what exactly constitutes an armed attack in cyberspace or the use of force—two critical elements of the international laws of armed conflict. Consequently, in the absence of much promise from either a legal or a technical perspective, states instead engage in that slow, laborious, and often frustrating practice of discussion and negotiation about agreed norms of what is and is not acceptable behavior (Carr 2017).

While the slow progress on this is frustrating for many, it is less surprising for those who study other issues of global security cooperation. There have been some important cyber security initiatives put forward by states that, when scrutinized, serve to highlight how the persistent injection of power politics has slowed down any potential progress on international cooperation. In fact, while cyber security is increasingly recognized as a serious threat to national and global security, it is rarely the end goal. Rather, the promotion of a broader set
of interests and goals specific to powerful states provides the context within which cyber security is navigated and negotiated. Put simply, some international cyber security proposals and initiatives have been subsumed by broader political agendas rather than evaluated on the basis of the quality of the content or ideas behind them. Given the extent of the growing implications for cyber insecurity, this could be regarded as a failure of states to adequately balance the perennial forces of great power politics with the pursuit of common or shared security goals.

**Conclusion**

In essence then, we come full circle back to an analysis of how political factors interact with technological developments—in a multidimensional way. As with any historical analysis, the purpose is to develop a better understanding of the context in which we move forward. The problems of cyber security are not simply technological. Rather, they are embedded in the ways that we choose to integrate that technology into a whole range of social, legal, economic, military, and political practices and processes. They are shaped by us and shape us. They are perceived differently in some ways and consistently in others. They cannot be resolved by technology or law alone because they spill out well beyond the borders of either of those domains.

One thing that is clear, however, is that, although politics may certainly be an impediment to better cyber security on one level, it will never be separated from it. Technological change responds to the needs of powerful actors—both state and non-state. It rarely develops in a context completely disconnected from perceived ‘problem solving’ (or threats) and how those problems are perceived and promoted is a function of power dynamics—within societies, within states, and within the international system. In order to ensure that the great opportunities of the third and fourth industrial revolutions are distributed in line with principles of equity, justice, and human rights, it is essential that we develop much more awareness of the political dynamics behind these movements. Doing so may allow us to deliver outcomes more conducive to global prosperity and harmony than those of the first and second industrial revolutions that have arguably institutionalized an unprecedented concentration of wealth as well as extraordinary benefits.

**Notes**

1. Abbate, Inventing the Internet, 40.